

Protein Sequence Searches - February 2005

All of the sequence databases on ABSS have recently been updated.

- Please note that the curators of the UniProt database have purged some temporary accession numbers from the most recent version of UniProt. These sequences have been assigned new permanent accession numbers. The new UniProt record may not contain the previous temporary accession number.
- If you encounter an accession number from an older search run against UniProt (results file extension **.rup**) that can no longer be found in the database, the permanent record with the new accession number can be found by searching the old accession number in the UniProt Protein Archive database (UniPARC) at:

<http://www.pir.uniprot.org/database/archive.shtml>

If you have any questions regarding this information or your results, please contact any STIC searcher.

When submitting sequence search results for scanning into IFW, please include a copy of this attachment to assist any future Examiners or members of the public who may encounter UniProt temporary accession numbers.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2005, 12:59:21 ; Search time 58.94 Seconds
(without alignments)
45.934 Million cell updates/sec

Title: SEQ1
Perfect score: 34
Sequence: 1 sgtsgsq 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Databases :
- 1: A_Geneseq_16Dec04:*
 - 2: Geneseq1980s:*
 - 3: Geneseq2000s:*
 - 4: Geneseq2001s:*
 - 5: Geneseq2002s:*
 - 6: Geneseq2003as:*
 - 7: Geneseq2003bs:*
 - 8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	100.0	7	3	AAU41588 Synthetic
2	34	100.0	253	3	ABM38107 Murine an
3	34	100.0	516	6	ABU48562 Protein e
4	34	100.0	1046	8	ADN18894 Bacterial
5	34	100.0	1629	4	ABM47521 Drosophil
6	31	91.2	80	4	AAU48718 Propionib
7	31	91.2	80	6	ABM45237 Propionib
8	31	91.2	81	3	AAU48718 Propionib
9	31	91.2	81	3	AAU48718 Propionib
10	31	91.2	81	3	AAU48718 Propionib
11	31	91.2	81	3	AAU48718 Propionib
12	31	91.2	81	3	AAU48718 Propionib
13	31	91.2	81	3	AAU48718 Propionib
14	31	91.2	107	6	AAE37596 HIV-1 non
15	31	91.2	107	6	AAE37596 HIV-1 non
16	31	91.2	252	3	AAU48718 Propionib
17	31	91.2	253	3	AAU48718 Propionib
18	31	91.2	426	8	ADN1273 Bacterial
19	31	91.2	620	5	ABP65371 Bifidobac
20	31	91.2	858	5	AAU48718 Propionib
21	31	91.2	1091	5	ABM48258 Listeria
22	31	91.2	2341	7	ADK11435 Drosophil
23	31	91.2	2342	4	ABM48258 Listeria
24	30	88.2	49	2	AAW38300 Subtilisi
25	30	88.2	49	2	AAW22749 Subtilisi

26	30	88.2	79	4	AAU41588	AAU41588 Propionib
27	30	88.2	79	6	ABM38107	ABM38107 Propionib
28	30	88.2	134	8	ABM80689	ABM80689 Tumour-as
29	30	88.2	157	4	AAU51002	AAU51002 Propionib
30	30	88.2	157	6	ABM47521	ABM47521 Propionib
31	30	88.2	185	7	AAO30818	AAO30818 Human cel
32	30	88.2	186	4	AAU51943	AAU51943 Propionib
33	30	88.2	186	6	ABM48462	ABM48462 Propionib
34	30	88.2	198	8	ADN14036	ADN14036 Human pro
35	30	88.2	274	1	AAU4862	AAU4862 Subtilisi
36	30	88.2	274	2	AAU4862	AAU4862 Subtilisi
37	30	88.2	274	2	AAU4862	AAU4862 Subtilisi
38	30	88.2	274	2	AAU4862	AAU4862 Subtilisi
39	30	88.2	274	2	AAU4862	AAU4862 Subtilisi
40	30	88.2	274	2	AAU4862	AAU4862 Subtilisi
41	30	88.2	274	2	AAU4862	AAU4862 Subtilisi
42	30	88.2	274	2	AAU4862	AAU4862 Subtilisi
43	30	88.2	274	2	AAU4862	AAU4862 Subtilisi
44	30	88.2	274	2	AAU4862	AAU4862 Subtilisi
45	30	88.2	274	2	AAU4862	AAU4862 Subtilisi

ALIGNMENTS

RESULT 1
AAU41588
ID AAY71086 standard; peptide; 7 AA.
XX AAY71086;
AC AAY71086;
XX AAY71086;
DT 21-SEP-2000 (first entry)
XX AAY71086;
DE Synthetic linker peptide #1 encoded by MV01JA oligonucleotide linker.
XX AAY71086;
KW llama; HC-V; heavy chain variable domain; antigen binding protein;
KW linker; conformational flexibility; multivalent binding protein; bi-head;
KW human chorionic gonadotropin; hCG; immunoassay; agglutination assay;
purification.
XX AAY71086;
OS Synthetic.
XX AAY71086;
FH Peptide
FT Peptide
FT Peptide
FT Peptide
FT Peptide
XX WO200024884-A2.
XX 04-MAY-2000.
XX 22-OCT-1999; 99WO-EP008323.
XX 27-OCT-1998; 98WO-EP006991.
XX 22-APR-1999; 99EP-00303118.
(UNIL) UNILEVER PLC.
(UNIL) UNILEVER NV.
(HIND-) HINDUSTAN LEVER LTD.
Frenken LGJ, Howell S, Van Der Vaart JM;
WPI: 2000-350728/30.
N-PSDB; AAD00657.
Use of a linker whose amino acid sequence confers restricted
conformational flexibility to generate multivalent and multispecific
antigen binding proteins.
Example 1.1d; Page 20; 50pp; English.
The present sequence is the synthetic linker peptide #1, encoded by the

CC oligonucleotide linker fragment, MV01JA. It consists of the last residue
 CC of the N-terminal HC-V fragment (S) and the first residue of the C-
 CC terminal HC-V fragment (Q), intersected by the connecting linker peptide.
 CC It is used for the construction of Saccharomyces cerevisiae episomal
 CC expression plasmid, pUR5330, encoding anti-hCG-anti-RR6 bispecific
 CC biheads, containing the linker peptide. The peptide linker confers
 CC restricted conformational flexibility for linking binding units in a
 CC multivalent binding protein. The linker is used to generate multivalent
 CC or multispecific antigen binding proteins for immunoassays, agglutination
 CC assays or for purification
 XX
 SQ Sequence 7 AA;

Query Match 100.0%; Score 34; DB 3; Length 7;
 Best Local Similarity 100.0%; Pred. No. 1.8e+06; Mismatches 0; Indels 0; Gaps 0;
 Matches 7; Conservative 0;

QY 1 SGTSGSQ 7
 |||||
 Db 1 SGTSGSQ 7

RESULT 2
 AAB28159
 ID AAB28159 standard; protein; 253 AA.
 XX
 AC AAB28159;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Murine anti-EGP-2 single chain Fv fragment.
 XX
 DE Murine; immunoglobulin; antigen-binding; framework region; carcinoma;
 KW c-erbB2; carcinoma-associated antigen.
 KW
 XX Mus sp.

XX WO200061635-A2.
 XX 19-OCT-2000.
 XX 10-APR-2000; 2000WO-EP003176.
 XX 09-APR-1999; 99EP-00107030.
 XX (UYZU-) UNIV ZUERICH.
 PA (PLUE/) PLUECKTHUN A.
 XX Plueckthun A, Honegger A, Willuda J;
 XX WPI; 2000-679468/66.
 XX
 XX Stabilizing chimeric immunoglobulin (Ig) involves setting up a stabilized
 PT antigen binding Ig or its fragment by replacing one or more residues
 PT present in acceptor Ig by those residues present in donor Ig.
 XX
 XX Claim 6; Page 51; 51pp; English.

XX The present invention relates to a method for stabilising a chimeric
 CC immunoglobulin (Ig). The method comprises identifying antigen-binding
 CC groups derived from donor Ig and framework regions derived from an
 CC acceptor Ig. The present sequence is one such donor Ig fragment. One or
 CC more of the residues present at the positions in the present sequence are
 CC replaced by those present at the corresponding positions in the donor Ig,
 CC after comparing the structural features of the VH domains of the acceptor
 CC Ig and the donor Ig. The acceptor Ig fragment used in the present
 CC invention is human anti-c-ErbB2 single chain Fv fragment 4D5 (AAB28158).
 CC The method of the present invention is useful for producing a
 CC pharmaceutical composition which can be used for treating human
 CC carcinomas, since c-erbB2 is a carcinoma-associated antigen
 XX
 XX Sequence 253 AA;

Query Match 100.0%; Score 34; DB 3; Length 253;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02; Mismatches 0; Indels 0; Gaps 0;
 Matches 7; Conservative 0;

QY 1 SGTSGSQ 7
 |||||
 Db 132 SGTSGSQ 138

RESULT 3
 ABU48562
 ID ABU48562 standard; protein; 516 AA.

XX AC ABU48562;
 XX
 DT 19-JUN-2003 (first entry)
 XX
 DE Protein encoded by Prokaryotic essential gene #34089.
 XX
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX
 OS Treponema pallidum.

XX WO200277183-A2.
 XX 03-OCT-2002.
 XX 21-MAR-2002; 2002WO-US009107.
 XX 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.
 XX Wang L, Zamudio C, Malone C, Haeelbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX WPI; 2003-029926/02.
 DR N-PSDB; ACA52432.
 DR
 XX New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids, required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 PS Claim 25; SEQ ID NO 76486; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational


```
XX SQ Sequence 1629 AA;
Query Match 100.0%; Score 34; DB 4; Length 1629;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGTSGSQ 7
    |||||
Db 404 SGTSGSQ 410

RESULT 6
AAU48718
ID AAU48718 standard; protein; 80 AA.
XX
AC AAU48718;
XX
DT 27-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #9614.
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
PN WO200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US012865.
XX
PR 21-APR-2000; 2000US-0199047P.
PR 02-JUN-2000; 2000US-0208941P.
PR 07-JUL-2000; 2000US-0216747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
WPI: 2001-616774/71.
DR N-PSDB; AAS59543.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris.
XX
PS Example 1; SEQ ID NO 9913; 1069pp; English.
XX
CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
```

```
XX SQ Sequence 80 AA;
Query Match 91.2%; Score 31; DB 4; Length 80;
Best Local Similarity 85.7%; Pred. No. 2.1e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGTSGSQ 7
    |||||
Db 29 SGTSGAQ 35

RESULT 7
ABM45237
ID ABM45237 standard; protein; 80 AA.
XX
AC ABM45237;
XX
DT 20-OCT-2003 (first entry)
XX
DE Propionibacterium acnes predicted ORF-encoded polypeptide #9913.
XX
KW Acne vulgaris; antisborrheic; dermatological; antibacterial;
KW immunostimulant; immune response; vaccine.
XX
OS Propionibacterium acnes.
XX
PN WO2003033515-A1.
XX
PD 24-APR-2003.
XX
PF 11-OCT-2002; 2002WO-US032727.
XX
PR 15-OCT-2001; 2001US-00978825.
XX
PA (CORI-) CORIXA CORP.
XX
PI Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;
PI Zhang Y, Wang S, Jen S, Lodes WJ, Benson DR, Jones R, Carter D;
PI Barth B, Vallieve-Douglass J;
XX
WPI: 2003-381789/36.
DR N-PSDB; ACF64472.
XX
PT New Propionibacterium acnes polypeptides and polynucleotides encoding the
PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
PT or for stimulating an immune response specific for a P. acnes protein.
XX
PS Example 1; SEQ ID NO 9913; 1481pp; English.
XX
CC The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
CC encoding a Propionibacterium acnes protein. The invention also relates to
CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
CC immunogenic fragments of P. acnes polypeptides. The invention
CC additionally encompasses expression vectors and host cells comprising a
CC polynucleotide of the invention; antibodies against polypeptides of the
CC invention; fusion proteins comprising a polypeptide of the invention; a
CC method for stimulating an immune response specific for a P. acnes
CC polypeptide and an isolated T cell population comprising T cells prepared
CC via this method; a vaccine composition (comprising P. acnes polypeptides,
CC polynucleotides, antibodies, fusion proteins, T cell populations, or
CC antigen-presenting cells that express the polypeptide); a method and kit
CC for detecting or determining the presence or absence of P. acnes in a
CC patient; and a method for inhibiting the development of P. acnes in a
CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
CC proteins, T cell populations or antigen-presenting cells that express the
CC polypeptides are useful for diagnosing, preventing or treating acne
CC vulgaris, or for stimulating an immune response specific for a P. acnes
CC protein. The polynucleotides can also be used as probes or primers for
CC nucleic acid hybridisation. The vaccine composition is useful for the
CC stimulation of an immune response against P. acnes, or for treating acne,
CC and the kit is useful for performing a diagnostic assay. The present
CC sequence represents a polypeptide predicted to be encoded by an ORF (open
```

CC reading frame) contained within the P. acnes polynucleotides of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 80 AA;

Query Match 91.2%; Score 31; DB 6; Length 80;

Best Local Similarity 85.7%; Pred. No. 2.1e+02; Indels 0; Gaps 0;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGTSGSQ 7

|||||:

Db 29 SGTSGAQ 35

RESULT 8

AAB69234

ID AAB69234 standard; protein; 81 AA.

XX AC AAB69234;

XX 12-SEP-2003 (revised)

DT 20-APR-2001 (first entry)

DE HIV-1 non-subtype B clone UC268A2 rev gene peptide.

XX HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;

KW vif; vpr; tat; rev; nef; vaccine.

XX Human immunodeficiency virus 1.

OS WO200026416-A1.

PN 11-MAY-2000.

PF 25-OCT-1999; 99WO-US024837.

PR 02-NOV-1998; 98US-00184418.

XX (UABR-) UAB RES FOUND.

PI Hahn BH, Shaw GM, Gao F;

XX WPI; 2000-365651/31.

XX Novel genomic nucleic acids of non-subtype B human immunodeficiency virus

PT type 1 useful for detecting and treating AIDS comprises a specific

PT nucleotide sequence.

XX Example 7; Fig 7; 131pp; English.

XX The present in invention provides the protein and coding sequences for a

CC number of human immunodeficiency virus (HIV) type 1 non-subtype B

CC isolates. The sequences shown include the near full-length coding

CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,

CC rev and nef proteins. These can be used to detect the presence of HIV-1

CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.

CC These antibodies can be used in vaccines to prevent and treat HIV

CC infection. (Updated on 12-SEP-2003 to standardise OS field)

XX Sequence 81 AA;

Query Match 91.2%; Score 31; DB 3; Length 81;

Best Local Similarity 85.7%; Pred. No. 2.1e+02; Indels 0; Gaps 0;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGTSGSQ 7

|||||:

Db 63 SGTSGTQ 69

RESULT 9

AAB69236

ID AAB69236 standard; protein; 81 AA.

XX AC AAB69236;

XX 12-SEP-2003 (revised)

DT 20-APR-2001 (first entry)

DE HIV-1 non-subtype B clone ZAM18A rev gene peptide.

XX HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;

KW vif; vpr; tat; rev; nef; vaccine.

XX Human immunodeficiency virus 1.

OS WO200026416-A1.

PN 11-MAY-2000.

PF 25-OCT-1999; 99WO-US024837.

PR 02-NOV-1998; 98US-00184418.

XX (UABR-) UAB RES FOUND.

PI Hahn BH, Shaw GM, Gao F;

XX WPI; 2000-365651/31.

XX Novel genomic nucleic acids of non-subtype B human immunodeficiency virus

PT type 1 useful for detecting and treating AIDS comprises a specific

PT nucleotide sequence.

XX Example 7; Fig 7; 131pp; English.

XX The present in invention provides the protein and coding sequences for a

CC number of human immunodeficiency virus (HIV) type 1 non-subtype B

CC isolates. The sequences shown include the near full-length coding

CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,

CC rev and nef proteins. These can be used to detect the presence of HIV-1

CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.

CC These antibodies can be used in vaccines to prevent and treat HIV

CC infection. (Updated on 12-SEP-2003 to standardise OS field)

XX Sequence 81 AA;

Query Match 91.2%; Score 31; DB 3; Length 81;

Best Local Similarity 85.7%; Pred. No. 2.1e+02; Indels 0; Gaps 0;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGTSGSQ 7

|||||:

Db 63 SGTSGTQ 69

RESULT 10

AAB69236

ID AAB69236 standard; protein; 81 AA.

XX AC AAB69236;

XX 12-SEP-2003 (revised)

DT 20-APR-2001 (first entry)

DE HIV-1 non-subtype B clone ZAM18A rev gene peptide.

XX HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;

KW vif; vpr; tat; rev; nef; vaccine.

XX Human immunodeficiency virus 1.

OS WO200026416-A1.

PN 11-MAY-2000.

PF 25-OCT-1999; 99WO-US024837.

PR 02-NOV-1998; 98US-00184418.

XX (UABR-) UAB RES FOUND.

PI Hahn BH, Shaw GM, Gao F;

XX WPI; 2000-365651/31.

XX Novel genomic nucleic acids of non-subtype B human immunodeficiency virus

PT type 1 useful for detecting and treating AIDS comprises a specific

PT nucleotide sequence.

XX Example 7; Fig 7; 131pp; English.

XX The present in invention provides the protein and coding sequences for a

CC number of human immunodeficiency virus (HIV) type 1 non-subtype B

CC isolates. The sequences shown include the near full-length coding

CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,

CC rev and nef proteins. These can be used to detect the presence of HIV-1

CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.

CC These antibodies can be used in vaccines to prevent and treat HIV

CC infection. (Updated on 12-SEP-2003 to standardise OS field)

XX Sequence 81 AA;

Query Match 91.2%; Score 31; DB 3; Length 81;

Best Local Similarity 85.7%; Pred. No. 2.1e+02; Indels 0; Gaps 0;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGTSGSQ 7

|||||:

Db 63 SGTSGTQ 69

RESULT 11

AAB69237

ID AAB69237 standard; protein; 81 AA.

XX AC AAB69237;

XX 12-SEP-2003 (revised)

DT 20-APR-2001 (first entry)

DE HIV-1 non-subtype B clone ZAM18A rev gene peptide.

XX HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;

KW vif; vpr; tat; rev; nef; vaccine.

XX Human immunodeficiency virus 1.

OS WO200026416-A1.

PN 11-MAY-2000.

PF 25-OCT-1999; 99WO-US024837.

PR 02-NOV-1998; 98US-00184418.

XX (UABR-) UAB RES FOUND.

PI Hahn BH, Shaw GM, Gao F;

XX WPI; 2000-365651/31.

XX Novel genomic nucleic acids of non-subtype B human immunodeficiency virus

PT type 1 useful for detecting and treating AIDS comprises a specific

PT nucleotide sequence.

XX Example 7; Fig 7; 131pp; English.

XX The present in invention provides the protein and coding sequences for a

CC number of human immunodeficiency virus (HIV) type 1 non-subtype B

CC isolates. The sequences shown include the near full-length coding

CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,

CC rev and nef proteins. These can be used to detect the presence of HIV-1

CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.

CC These antibodies can be used in vaccines to prevent and treat HIV

CC infection. (Updated on 12-SEP-2003 to standardise OS field)

XX Sequence 81 AA;

Query Match 91.2%; Score 31; DB 3; Length 81;

Best Local Similarity 85.7%; Pred. No. 2.1e+02; Indels 0; Gaps 0;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGTSGSQ 7

|||||:

Db 63 SGTSGTQ 69

RESULT 12

AAB69238

ID AAB69238 standard; protein; 81 AA.

XX AC AAB69238;

XX 12-SEP-2003 (revised)

DT 20-APR-2001 (first entry)

DE HIV-1 non-subtype B clone ZAM18A rev gene peptide.

XX HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;

KW vif; vpr; tat; rev; nef; vaccine.

XX Human immunodeficiency virus 1.

OS WO200026416-A1.

PN 11-MAY-2000.

PF 25-OCT-1999; 99WO-US024837.

PR 02-NOV-1998; 98US-00184418.

XX (UABR-) UAB RES FOUND.

PI Hahn BH, Shaw GM, Gao F;

XX WPI; 2000-365651/31.

XX Novel genomic nucleic acids of non-subtype B human immunodeficiency virus

PT type 1 useful for detecting and treating AIDS comprises a specific

PT nucleotide sequence.

XX Example 7; Fig 7; 131pp; English.

XX The present in invention provides the protein and coding sequences for a

CC number of human immunodeficiency virus (HIV) type 1 non-subtype B

CC isolates. The sequences shown include the near full-length coding

CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,

CC rev and nef proteins. These can be used to detect the presence of HIV-1

CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.

CC These antibodies can be used in vaccines to prevent and treat HIV

CC infection. (Updated on 12-SEP-2003 to standardise OS field)

XX Sequence 81 AA;

Query Match 91.2%; Score 31; DB 3; Length 81;

Best Local Similarity 85.7%; Pred. No. 2.1e+02; Indels 0; Gaps 0;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGTSGSQ 7

|||||:

Db 63 SGTSGTQ 69

RESULT 13

AAB69239

ID AAB69239 standard; protein; 81 AA.

XX AC AAB69239;

XX 12-SEP-2003 (revised)

DT 20-APR-2001 (first entry)

DE HIV-1 non-subtype B clone ZAM18A rev gene peptide.

XX HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;

KW vif; vpr; tat; rev; nef; vaccine.

XX Human immunodeficiency virus 1.

OS WO200026416-A1.

PN 11-MAY-2000.

PF 25-OCT-1999; 99WO-US024837.

PR 02-NOV-1998; 98US-00184418.

XX (UABR-) UAB RES FOUND.

PI Hahn BH, Shaw GM, Gao F;

XX WPI; 2000-365651/31.

XX Novel genomic nucleic acids of non-subtype B human immunodeficiency virus

PT type 1 useful for detecting and treating AIDS comprises a specific

PT nucleotide sequence.

XX Example 7; Fig 7; 131pp; English.

XX The present in invention provides the protein and coding sequences for a

CC number of human immunodeficiency virus (HIV) type 1 non-subtype B

CC isolates. The sequences shown include the near full-length coding

CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,

CC rev and nef proteins. These can be used to detect the presence of HIV-1

CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.

CC These antibodies can be used in vaccines to prevent and treat HIV

CC infection. (Updated on 12-SEP-2003 to standardise OS field)

XX Sequence 81 AA;

Query Match 91.2%; Score 31; DB 3; Length 81;

Best Local Similarity 85.7%; Pred. No. 2.1e+02; Indels 0; Gaps 0;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGTSGSQ 7

|||||:

Db 63 SGTSGTQ 69

RESULT 14

AAB69240

ID AAB69240 standard; protein; 81 AA.

XX AC AAB69240;

XX 12-SEP-2003 (revised)

DT 20-APR-2001 (first entry)

DE HIV-1 non-subtype B clone ZAM18A rev gene peptide.

XX HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;

KW vif; vpr; tat; rev; nef; vaccine.

XX Human immunodeficiency virus 1.

OS WO200026416-A1.

PN 11-MAY-2000.

PF 25-OCT-1999; 99WO-US024837.

PR 02-NOV-1998; 98US-00184418.

XX (UABR-) UAB RES FOUND.

PI Hahn BH, Shaw GM, Gao F;

XX WPI; 2000-365651/31.

XX Novel genomic nucleic acids of non-subtype B human immunodeficiency virus

PT type 1 useful for detecting and treating AIDS comprises a specific

PT nucleotide sequence.

XX Example 7; Fig 7; 131pp; English.

XX The present in invention provides the protein and coding sequences for a

CC number of human immunodeficiency virus (HIV) type 1 non-subtype B

CC isolates. The sequences shown include the near full-length coding

CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,

CC rev and nef proteins. These can be used to detect the presence of HIV-1

CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.

CC These antibodies can be used in vaccines to prevent and treat HIV

CC infection. (Updated on 12-SEP-2003 to standardise OS field)

XX Sequence 81 AA;

Query Match 91.2%; Score 31; DB 3; Length 81;

Best Local Similarity


```

Db      63 SGTSGTQ 69
|||||:|
RESULT 13
AAB69249
ID AAB69249 standard; protein; 97 AA.
XX
AC AAB69249;
XX
DT 12-SEP-2003 (revised)
DT 20-APR-2001 (first entry)
XX
DE HIV-1 non-subtype B rev gene consensus peptide #7.
XX
KW HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;
KW vif; vpr; tat; rev; nef; vaccine.
XX
OS Human immunodeficiency virus 1.
XX
PN WO200026416-A1.
XX
PD 11-MAY-2000.
XX
PF 25-OCT-1999; 99WO-US024837.
XX
PR 02-NOV-1998; 98US-00184418.
XX
PA (UABR-) UAB RES FOUND.
XX
PI Hahn BH, Shaw GM, Gao F;
XX
DR WPI; 2000-365651/31.
XX
PT Novel genomic nucleic acids of non-subtype B human immunodeficiency virus
PT type 1 useful for detecting and treating AIDS comprises a specific
PT nucleotide sequence.
XX
PS Example 7; Fig 7; 131pp; English.
XX
CC The present invention provides the protein and coding sequences for a
CC number of human immunodeficiency virus (HIV) type 1 non-subtype B
CC isolates. The sequences shown include the near full-length coding
CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,
CC rev and nef proteins. These can be used to detect the presence of HIV-1
CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.
CC These antibodies can be used in vaccines to prevent and treat HIV
CC infection. (Updated on 12-SEP-2003 to standardise OS field)
XX
SQ Sequence 97 AA;
Query Match 91.2%; Score 31; DB 3; Length 97;
Best Local Similarity 85.7%; Pred. No. 2.6e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 SGTSGSQ 7
Db 63 SGTSGTQ 69
|||||:|
RESULT 14
AAB37596
ID AAB37596 standard; protein; 107 AA.
XX
AC AAB37596;
XX
DT 23-OCT-2003 (revised)
DT 27-AUG-2003 (first entry)
XX
DE HIV-1 subtype C rev consensus protein fragment.
XX
KW Regulatory gene; accessory gene; HIV; human immunodeficiency virus;
KW vaccine; infection; gene therapy; rev.

```

```

XX OS Human immunodeficiency virus 1.
XX PN WO2003037919-A2.
XX PD 08-MAY-2003.
XX PF 31-OCT-2002; 2002WO-IB004550.
XX PR 31-OCT-2001; 2001ZA-00008978.
XX PA (SAME-) SOUTH AFRICAN MEDICAL RES COUNCIL.
XX PA (UYCA-) UNIV CAPE TOWN.
XX PI Williamson C, Van Harmelen JH, Gray CM, Bourn W, Karim SA;
XX DR WPI; 2003-430497/40.
XX PT New molecules comprising HIV-1 subtype isolate regulatory/accessory
XX PT genes, useful for manufacturing a vaccine for treating or preventing HIV
XX PT infection.
XX PS Claim 22; Page 36; 97pp; English.
XX CC The invention relates to molecules comprising HIV-1 subtype isolate
XX CC regulatory/accessory genes (tat, nef and rev genes) and modifications and
XX CC derivatives thereof. The invention also provides proteins encoded by such
XX CC genes. Sequences of the invention are useful for manufacturing vaccines
XX CC for treating or preventing human immunodeficiency virus (HIV) infections.
XX CC They are also useful in gene therapy. The present sequence is HIV-1
XX CC subtype C rev consensus protein fragment. (Updated on 23-OCT-2003 to
XX CC standardise OS field)
XX SQ Sequence 107 AA;
Query Match 91.2%; Score 31; DB 6; Length 107;
Best Local Similarity 85.7%; Pred. No. 2.8e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 SGTSGSQ 7
Db 89 SGTSGTQ 95
|||||:|
RESULT 15
AAB37590
ID AAB37590 standard; protein; 107 AA.
XX
AC AAB37590;
XX
DT 23-OCT-2003 (revised)
DT 27-AUG-2003 (first entry)
XX
DE HIV-1 subtype C isolate Du422 rev protein.
XX
KW Regulatory gene; accessory gene; HIV; human immunodeficiency virus;
KW vaccine; infection; gene therapy; rev.
XX
OS Human immunodeficiency virus 1.
XX PN WO2003037919-A2.
XX PD 08-MAY-2003.
XX PF 31-OCT-2002; 2002WO-IB004550.
XX PR 31-OCT-2001; 2001ZA-00008978.
XX PA (SAME-) SOUTH AFRICAN MEDICAL RES COUNCIL.
XX PA (UYCA-) UNIV CAPE TOWN.
XX PI Williamson C, Van Harmelen JH, Gray CM, Bourn W, Karim SA;
XX

```

DR WPI; 2003-430497/40.
DR N-PSDB; AAD37590.
XX
PT New molecules comprising HIV-1 subtype isolate regulatory/accessory
PT genes, useful for manufacturing a vaccine for treating or preventing HIV
PT infection.
XX
XX Claim 17; Fig 8; 97pp; English.
XX
CC The invention relates to molecules comprising HIV-1 subtype isolate
CC regulatory/accessory genes (tat, nef and rev genes) and modifications and
CC derivatives thereof. The invention also provides proteins encoded by such
CC genes. Sequences of the invention are useful for manufacturing vaccines
CC for treating or preventing human immunodeficiency virus (HIV) infections.
CC They are also useful in gene therapy. The present sequence is HIV-1
CC subtype C isolate Du422 rev protein. (Updated on 23-OCT-2003 to
CC standardise OS field)
XX
SQ Sequence 107 AA;
Query Match 91.2%; Score 31; DB 6; Length 107;
Best Local Similarity 85.7%; Pred. No. 2.8e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 SGTSGSQ 7
Db 89 SGTSGTQ 95
|||||:
Search completed: July 18, 2005, 13:40:24
Job time : 61.94 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2005, 12:59:21 ; Search time 53.06 Seconds
(without alignments)
67.557 Million cell updates/sec

Title: SEQ1
Perfect score: 34
Sequence: 1 sgtsqsg 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Uniprot 03: *
1: uniprot_sprot: *
2: uniprot_trembl: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	34	100.0	516	1 Y369 TREPA	O83384 treponema p
2	34	100.0	1046	2 Q07653	Q07653 saccharomyc
3	34	100.0	1427	2 Q8G122	Q8G122 mycoplasma
4	34	100.0	1669	1 DMBD DROME	Q9v4w3 drosophila
5	34	100.0	2277	2 Q9U0G5	Q9U0G5 plasmodium
6	31	91.2	90	2 Q74901	Q74901 human immun
7	31	91.2	90	2 Q74903	Q74903 human immun
8	31	91.2	90	2 Q74904	Q74904 human immun
9	31	91.2	91	2 Q74902	Q74902 human immun
10	31	91.2	95	2 Q8AC75	Q8AC75 human immun
11	31	91.2	100	2 Q8UT12	Q8UT12 human immun
12	31	91.2	100	2 Q6S729	Q6S729 human immun
13	31	91.2	101	2 Q8AU61	Q8AU61 human immun
14	31	91.2	107	2 Q8UT21	Q8UT21 human immun
15	31	91.2	107	2 Q8UT48	Q8UT48 human immun
16	31	91.2	107	2 Q8UT57	Q8UT57 human immun
17	31	91.2	107	2 Q8UT66	Q8UT66 human immun
18	31	91.2	107	2 Q8UT75	Q8UT75 human immun
19	31	91.2	107	2 Q8UT84	Q8UT84 human immun
20	31	91.2	107	2 Q8UT93	Q8UT93 human immun
21	31	91.2	107	2 Q8UT13	Q8UT13 human immun
22	31	91.2	107	2 Q8UTN7	Q8UTN7 human immun
23	31	91.2	107	2 Q8UTP6	Q8UTP6 human immun
24	31	91.2	107	2 Q8UTR4	Q8UTR4 human immun
25	31	91.2	107	2 Q8UT12	Q8UT12 human immun
26	31	91.2	107	2 Q8UTU1	Q8UTU1 human immun
27	31	91.2	107	2 Q900M0	Q900M0 human immun
28	31	91.2	107	2 Q900M3	Q900M3 human immun
29	31	91.2	107	2 Q900M6	Q900M6 human immun
30	31	91.2	107	2 Q900M7	Q900M7 human immun
31	31	91.2	107	2 Q901X4	Q901X4 human immun

32	31	91.2	107	2	Q901Y3	Q901Y3 human immun
33	31	91.2	107	2	Q901Z2	Q901Z2 human immun
34	31	91.2	107	2	Q994F9	Q994F9 human immun
35	31	91.2	107	2	Q994N1	Q994N1 human immun
36	31	91.2	107	2	Q66TQ7	Q66TQ7 human immun
37	31	91.2	107	2	Q66TT6	Q66TT6 human immun
38	31	91.2	107	2	Q6PQZ6	Q6PQZ6 human immun
39	31	91.2	107	2	Q6S7T2	Q6S7T2 human immun
40	31	91.2	107	2	Q6S7U9	Q6S7U9 human immun
41	31	91.2	107	2	Q6S7Y3	Q6S7Y3 human immun
42	31	91.2	107	2	Q6S815	Q6S815 human immun
43	31	91.2	107	2	Q6S831	Q6S831 human immun
44	31	91.2	107	2	Q6S846	Q6S846 human immun
45	31	91.2	107	2	Q6S878	Q6S878 human immun

ALIGNMENTS

RESULT 1
Y369 TREPA ID Y369 TREPA STANDARD; PRT; 516 AA.
AC O83384;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Hypothetical protein TP0369 precursor.
GN OrderedLocustNames=TP0369;
OS Treponema pallidum.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Nichols;
RX MEDLINE=98332770; PubMed=9665876; DOI=10.1126/science.281.5375.375;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R.J., Winn M.L., Hickey E.K., Clayton R.A., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S.L., Peterson J.D.,
RA Khalak H.G., Richardson D.L., Howell J.K., Chidambaram M.,
RA Uterback T.R., McDonald L.A., Artach P., Bowman C., Cotton M.D.,
RA Fujii C., Garland S.A., Hatch B., Horst K., Roberts K.M., Sandusky M.,
RA Weidman J.F., Smith H.O., Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
spirochete."
RL Science 281:375-388(1998).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; AE001216; AAC65360.1; --
DR PIR; H71332; H71332.
DR TIGR; TP0369; --
DR InterPro; IPR008941; TPR-like.
KW Complete proteome; Hypothetical protein; Signal.
FT SIGNAL 1 17 Potential.
FT CHAIN 18 516 Hypothetical protein TP0369.
SQ SEQUENCE 516 AA; 55859 MW; 1FBC37CE5A0F0CCF CRC64;

Query Match 100.0%; Score 34; DB 1; Length 516;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SGTSGSQ 7

Db 323 SGTSGSQ 329

RESULT 2

Q07653
 ID Q07653 PRELIMINARY; PRT; 1046 AA.
 AC Q07653
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-WAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE S.cerevisiae chromosome IV reading frame ORF YDL223c.
 GN Names:HBT1;
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Rasmussen S.W.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MIPS;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; Z74271; CAA98802.1; --
 DR PIR; S67786; S67786.
 DR SGD; S000002382; HBT1.
 DR GO; GO:0005937; C:shmoo tip; IDA.
 DR GO; GO:000753; P:cellular morphogenesis during conjugation w. . .; IMP.
 DR InterPro; IPR000102; MAP1B_neuraxin.
 DR PROSITE; PS00230; MAP1B_NEURAXIN; UNKNOWN 1.
 SQ SEQUENCE 1046 AA; 113616 MW; DC5E3EB5DDAF4A05 CRC64;

Query Match 100.0%; Score 34; DB 2; Length 1046;
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGTSGSQ 7
 DB 513 SGTSGSQ 519

RESULT 3
 Q8G1Z2 PRELIMINARY; PRT; 1427 AA.
 ID Q8G1Z2
 AC Q8G1Z2
 DT 01-WAR-2003 (TrEMBLrel. 23, Created)
 DT 01-WAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-WAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Lipoprotein.
 GN Names:lpps;
 OS Mycoplasma conjugativae.
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=45361;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22562633; PubMed=12676664;
 RX DOI=10.1128/AEM.69.4.1913-1919.2003;
 RA Belloy L., Janovsky M., Vilgii E.M., Pilo P., Giacometti M., Frey J.;
 RT "Molecular epidemiology of Mycoplasma conjugativae in Caprinae: transmission across species in natural outbreaks.";
 RL Appl. Environ. Microbiol. 69:1913-1919(2003).
 DR EMBL; AU514404; CAD55813.1; --
 SQ SEQUENCE 1427 AA; 155548 MW; 9155221E1CA91B7C CRC64;

Query Match 100.0%; Score 34; DB 2; Length 1427;
 Best Local Similarity 100.0%; Pred. No. 4.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGTSGSQ 7
 DB 1197 SGTSGSQ 1203

RESULT 4
 DMDB_DROME

ID Q9VDM3; O62529; STANDARD; PRT; 1669 AA.
 AC Q9VDM3; O9BK99;
 DT 05-JUL-2004 (Rel. 44, Created)
 DT 05-JUL-2004 (Rel. 44, Last sequence update)
 DT 25-JAN-2005 (Rel. 46, Last annotation update)
 DE Dystrophin, isoform B.
 GN Name=Dys; ORFNames=CG31175;
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21125212; PubMed=11223239; DOI=10.1016/S0378-1119(00)00584-9;
 RA Neuman S., Kaban A., Volk T., Yaffe D., Nudel U.;
 RT "The dystrophin / utrophin homologues in Drosophila and in sea urchin.";
 RL Gene 263:17-29(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.M.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Fannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RN [3]
 RP GENOME REANNOTATION, AND ALTERNATIVE SPLICING.
 RX MEDLINE=22426069; PubMed=12537572;
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.O.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a systematic review.";

Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).

[4]

SEQUENCE OF 643-1598 FROM N.A.

MDLINE-98167843; PubMed=9499411; DOI=10.1093/hmg/7.4.589;

Roberts R.G., Bobrow M.;

"Dystrophins in vertebrates and invertebrates.";

Hum. Mol. Genet. 7:589-595(1998).

-1- FUNCTION: May play a role in anchoring the cytoskeleton to the plasma membrane.

-1- ALTERNATIVE PRODUCTS:

Event=Alternative splicing; Named isoforms=5;

Name=B;

IsoId=Q9VDW3-1; Sequence=Displayed;

Name=A; Synonyms=DLP2;

IsoId=Q9VDW6-1; Sequence=External;

Name=C;

IsoId=Q9VDW6-2; Sequence=External;

Name=DLP1;

IsoId=Q9VDW6-3; Sequence=External;

Name=DLP3;

IsoId=Q9VDW6-4; Sequence=External;

-1- SIMILARITY: Contains 4 spectrin repeats.

-1- SIMILARITY: Contains 1 WW domain.

-1- SIMILARITY: Contains 1 ZZ-type zinc finger.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AF300294; AAK15257.1; -

EMBL; AE003726; AAF55676.3; -

EMBL; X99757; CAA68088.1; -

HSSP; P11532; LEG3.

FlyBase; FBgn0024242; Dys.

InterPro; IPR010983; BF_Hand like.

InterPro; IPR002017; Spectrin.

InterPro; IPR001202; WW Rep5 WWP.

InterPro; IPR000433; ZnF_ZZ.

Pfam; PF00435; Spectrin; 5.

Pfam; PF00397; WW; 1.

Pfam; PF00569; ZZ; 1.

SMART; SM00150; SPEC; 4.

SMART; SM00456; WW; 1.

SMART; SM00291; ZnF_ZZ; 1.

PROSITE; PS01159; WW_DOMAIN_1; FALSE_NEG.

PROSITE; PS50020; WW_DOMAIN_2; 1.

PROSITE; PS01357; ZF_ZZ_1; FALSE_NEG.

PROSITE; PS50135; ZF_ZZ_2; 1.

KW Actin-binding; Alternative splicing; Calcium-binding; Cytoskeleton;

KW Repeat; Structural protein; Zinc-finger.

FT DOMAIN 541 643

FT DOMAIN 650 747

FT DOMAIN 754 883

FT DOMAIN 890 990

FT DOMAIN 1021 1054

FT ZN_FING 1278 1325

FT CONFLICT 643 644

FT CONFLICT 699 699

FT CONFLICT 734 736

FT CONFLICT 742 742

FT CONFLICT 772 772

FT CONFLICT 783 784

FT CONFLICT 787 787

FT CONFLICT 854 854

FT CONFLICT 872 872

FT CONFLICT 886 886

FT CONFLICT 894 894

FT CONFLICT 923 930

FT CONFLICT 957 957

Q -> E (in Ref. 4).

FT CONFLICT 990 1000

FT CONFLICT 1045 1045

FT CONFLICT 1102 1106

FT CONFLICT 1110 1110

FT CONFLICT 1130 1130

FT CONFLICT 1294 1294

FT CONFLICT 1598 1598

SQ SEQUENCE 1669 AA; 185310 MW; B264332783162291 CRC64;

Query Match 100.0%; Score 34; DB 1; Length 1669;

Best Local Similarity 100.0%; Pred. No. 4.9e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGTSGSQ 7

Db 404 SGTSGSQ 410

RESULT 5

Q900G5 PRELIMINARY; PRT; 2277 AA.

AC Q900G5;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Erythrocyte membrane protein 1 (PFEMPL).

GN Names=PF0625c; Synonyms=VAR;

OS Plasmodium falciparum (isolate 3D7).

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI_TaxID=36329;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=2255708; PubMed=1236867; DOI=10.1038/nature01095;

RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D., Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K., Buckee C.O., Burrows C., Cherevach I., Chillingworth C., Corton C., Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C., Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J., Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z., Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P., Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A., Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N., Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L., Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E., Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M., Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K., Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J., Sulston J.E., Craig A., Newbold C., Barrell B.G.;

RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";

RL Nature 419:527-531(2002).

RN [2]

RP SEQUENCE FROM N.A.

RA Devlin K., Pain A., Berriman B., Hall N., Bowman S., Churcher C., Harris B., Harris D., Lawson D., Quail M., Barrell B.;

RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL035475; CAB62899.1; -

DR GO; GO:0005539; F:glycosaminoglycan binding; IEA.

DR GO; GO:0009405; P:pathogenesis; IEA.

DR InterPro; IPR004258; PFEMP.

DR Pfam; PF03011; PFEMP; 2.

SQ SEQUENCE 2277 AA; 256122 MW; E4262CCA69DDEF93 CRC64;

Query Match 100.0%; Score 34; DB 2; Length 2277;

Best Local Similarity 100.0%; Pred. No. 7e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGTSGSQ 7

Db 610 SGTSGSQ 616

RESULT 6

Q74901


```

DR GO: GO:0003700; F:transcription factor activity; IEA.
DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro: IPR000625; REV_protein.
DR Pfam: PF00424; REV; 1.
FT NON_TER 1 1
FT NON_TER 91 91
SQ SEQUENCE 91 AA; 10257 MW; 879ABD5DBD8EF5E5 CRC64;

Query Match          91.2%; Score 31; DB 2; Length 91;
Best Local Similarity 85.7%; Pred. No. 82;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGTSGSQ 7
Db 63 SGTSGTQ 69

RESULT 10
QBAC75 PRELIMINARY; PRT; 95 AA.
AC QBAC75;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Truncated rev protein.
GN Name=rev;
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IV002;
RX MEDLINE=22375630; PubMed=12487821; DOI=10.1089/08922202320886370;
RA zur Megede J., Engelbrecht S., De Oliveira T., Cassol S., Scriba T.J.,
RA van Rensburg E.J., Barnett S.W.;
RT "Novel evolutionary analyses of full-length HIV type 1 subtype C
RT molecular clones from Cape Town, South Africa.";
RL AIDS Res. Hum. Retroviruses 18:1327-1332(2002).
DR EMBL; AY162224; AAN75303.1;
DR GO: GO:0042025; C:host cell nucleus; IEA.
DR GO: GO:0003700; F:transcription factor activity; IEA.
DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro: IPR000625; REV_protein.
DR Pfam: PF00424; REV; 1.
SQ SEQUENCE 95 AA; 10815 MW; CADDR2B7F710F47C CRC64;

Query Match          91.2%; Score 31; DB 2; Length 95;
Best Local Similarity 85.7%; Pred. No. 86;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGTSGSQ 7
Db 89 SGTSGTQ 95

RESULT 11
QBUT12 PRELIMINARY; PRT; 100 AA.
AC QBUT12;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Rev protein.
GN Name=rev;
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21988475; PubMed=11991972;
RX DOI=10.1128/JVI.76.11.5435-5451.2002;
RA Novitsky V., Smith U.R., Gilbert P., McLane M.F., Chigwedere P.,
RA Williamson C., Ndung'u T., Klein I., Chang S.-Y., Peter T., Thior I.,

```

```

RA Foley B.T., Gaoekwe S., Rybak N., Gaseitsiwe S., Vannberg F.,
RA Marlink R., Lee T.-H., Essex M.;
RT "Human immunodeficiency virus type 1 subtype C molecular phylogeny:
RT consensus sequence for an AIDS vaccine design?";
RT J. Virol. 76:5435-5451(2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Novitsky V.A., McLane M.F., Chigwedere P., Ndung'u T., Klein I.,
RA Chang S.-Y., Peter T., Thior I., Rybak N., Gaseitsiwe S., Vannberg F.,
RA Marlink R., Lee T.-H., Essex M.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF43105; AAL34837.1;
DR GO: GO:0042025; C:host cell nucleus; IEA.
DR GO: GO:0003700; F:transcription factor activity; IEA.
DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR Pfam: PF00424; REV; 1.
SQ SEQUENCE 100 AA; 11310 MW; B1EA97EF4B19A504 CRC64;

Query Match          91.2%; Score 31; DB 2; Length 100;
Best Local Similarity 85.7%; Pred. No. 91;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGTSGSQ 7
Db 89 SGTSGTQ 95

RESULT 12
Q6S7Z9 PRELIMINARY; PRT; 100 AA.
ID Q6S7Z9;
AC Q6S7Z9;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Rev protein.
GN Name=rev;
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RG HIV HLA epitope mapping;
RA Korber B.T., Mullins J.I., Goulder P.J., Brander C., Kiepiela P.,
RA Moore S., Shindo N., Walker B.;
RL Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY463228; AAR27702.1;
DR GO: GO:0042025; C:host cell nucleus; IEA.
DR GO: GO:0003700; F:transcription factor activity; IEA.
DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
SQ SEQUENCE 100 AA; 11129 MW; 04ACBF2A87C7BA3B CRC64;

Query Match          91.2%; Score 31; DB 2; Length 100;
Best Local Similarity 85.7%; Pred. No. 91;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGTSGSQ 7
Db 89 SGTSGTQ 95

RESULT 13
QB6U61 PRELIMINARY; PRT; 101 AA.
ID QB6U61;
AC QB6U61;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Rev protein.
GN Name=rev;
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=22190101; PubMed=12201911; DOI=10.1089/0892220260190362;
RA Papathanasopoulos M.A., Cilliers T., Morris L., Mokili J.L.,
RA Dowling W., Birk D.L., McCutchan F.E.;
RT "Full-length genome analysis of HIV-1 subtype C utilizing CXCR4 and
RL intersubtype recombinants isolated in South Africa.";
RL AIDS Res. Hum. Retroviruses 18:879-886(2002).
DR EMBL; AF411967; AAN47132.1; -.
DR GO; GO:0042025; C:host cell nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR Interpro; IPR000625; REV_protein.
DR Pfam; PF00424; REV; 1.
SQ SEQUENCE 101 AA; 11402 MW; 4B1C61265DB07AE7 CRC64;

Query Match          91.2%; Score 31; DB 2; Length 101;
Best Local Similarity 85.7%; Pred. No. 92;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGTSGSQ 7
Db      |||||:|
      90 SGTSGTQ 96

RESULT 14
Q8UT21 PRELIMINARY; PRT; 107 AA.
AC Q8UT21;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Rev protein.
GN Name=rev;
OS Human immunodeficiency virus 1.
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21988475; PubMed=11991972;
RX DOI=10.1128/JVI.76.11.5435-5451.2002;
RA Novitsky V., Smith U.R., Gilbert P., McLane M.F., Chigwedere P.,
RA Williamson C., Ndung'u T., Klein I., Chang S.-Y., Peter T., Thior I.,
RA Foley B.T., Gaolekwe S., Rybak N., Gaseitsiwe S., Vannberg F.,
RA Marlink R., Lee T.-H., Essex M.;
RA "Human immunodeficiency virus type 1 subtype C molecular phylogeny:
RT consensus sequence for an AIDS vaccine design?";
RL J. Virol. 76:5435-5451(2002).
RN [2]
RP SEQUENCE FROM N.A.
RX Novitsky V.A., McLane M.F., Chigwedere P., Ndung'u T., Klein I.,
RA Chang S.-Y., Peter T., Thior I., Rybak N., Gaseitsiwe S., Vannberg F.,
RA Marlink R., Lee T.-H., Essex M.;
RA Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF443104; AAL34828.1; -.
DR GO; GO:0042025; C:host cell nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR Pfam; PF00424; REV; 1.
SQ SEQUENCE 107 AA; 11935 MW; F40117B087B5257A CRC64;

Query Match          91.2%; Score 31; DB 2; Length 107;
Best Local Similarity 85.7%; Pred. No. 99;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGTSGSQ 7
Db      |||||:|
      89 SGTSGTQ 95

RESULT 15
Q8UT48 PRELIMINARY; PRT; 107 AA.
AC Q8UT48;

```

```

DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Rev protein.
GN Name=rev;
OS Human immunodeficiency virus 1.
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21988475; PubMed=11991972;
RX DOI=10.1128/JVI.76.11.5435-5451.2002;
RA Novitsky V., Smith U.R., Gilbert P., McLane M.F., Chigwedere P.,
RA Williamson C., Ndung'u T., Klein I., Chang S.-Y., Peter T., Thior I.,
RA Foley B.T., Gaolekwe S., Rybak N., Gaseitsiwe S., Vannberg F.,
RA Marlink R., Lee T.-H., Essex M.;
RA "Human immunodeficiency virus type 1 subtype C molecular phylogeny:
RT consensus sequence for an AIDS vaccine design?";
RL J. Virol. 76:5435-5451(2002).
RN [2]
RP SEQUENCE FROM N.A.
RX Novitsky V.A., McLane M.F., Chigwedere P., Ndung'u T., Klein I.,
RA Chang S.-Y., Peter T., Thior I., Rybak N., Gaseitsiwe S., Vannberg F.,
RA Marlink R., Lee T.-H., Essex M.;
RA Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF443101; AAL34801.1; -.
DR GO; GO:0042025; C:host cell nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR Pfam; PF00424; REV; 1.
SQ SEQUENCE 107 AA; 11883 MW; F3C551DE133CG274 CRC64;

Query Match          91.2%; Score 31; DB 2; Length 107;
Best Local Similarity 85.7%; Pred. No. 99;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGTSGSQ 7
Db      |||||:|
      89 SGTSGTQ 95

Search completed: July 18, 2005, 13:33:16
Job time : 55.06 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2005, 13:09:46 ; Search time 16.94 Seconds
(without alignments)
62.478 Million cell updates/sec

Title: SEQ2
Perfect score: 45
Sequence: 1 sssasasaaq 11
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	88.9	239	2 S37923	cell wall protein
2	39	86.7	319	2 B86433	hypothetical prote
3	37	82.2	243	2 D84629	hypothetical prote
4	37	82.2	269	2 C86321	P6A14.11 protein -
5	37	82.2	349	2 T41394	hypothetical serin
6	37	82.2	354	2 S39406	homeotic protein o
7	37	82.2	364	2 I48188	gene NKx6.1 protei
8	37	82.2	600	2 S07638	spore coat protein
9	37	82.2	775	1 EDBE11	immediate-early pr
10	37	82.2	825	1 EDBEXD	immediate-early pr
11	37	82.2	1077	2 A44067	serine-rich protei
12	37	82.2	1198	2 T20262	hypothetical prote
13	37	82.2	1203	2 C89217	protein C55A6.2 [i
14	37	82.2	1283	2 T49804	hypothetical prote
15	36	80.0	217	2 B81067	conserved hypothet
16	36	80.0	217	2 C81801	probable lipoprote
17	36	80.0	268	2 C84585	hypothetical prote
18	36	80.0	323	2 G96544	hypothetical prote
19	36	80.0	461	2 JN0097	secreted 45K prote
20	36	80.0	470	2 T04356	isocitrate dehydro
21	36	80.0	482	2 T04355	isocitrate dehydro
22	36	80.0	743	2 T02828	conserved hypothet
23	36	80.0	1677	2 T43021	vitellogenin precu
24	35	77.8	232	1 A25108	homeotic protein H
25	35	77.8	592	1 S13391	endo-1,4-beta-xyla
26	35	77.8	779	2 A57177	NIMA-like protein
27	35	77.8	1111	2 T26972	hypothetical prote
28	35	77.8	2342	2 T13412	hypothetical prote
29	34	75.6	104	2 S49803	hypothetical prote

30	34	75.6	125	2 S69871	hypothetical prote
31	34	75.6	153	2 S67294	hypothetical prote
32	34	75.6	154	2 S55017	hypothetical prote
33	34	75.6	187	2 S69466	hypothetical prote
34	34	75.6	225	2 B84653	TINY-like AP2 doma
35	34	75.6	225	2 S59852	DNA-binding protei
36	34	75.6	229	2 JC7219	nuclear protein SR
37	34	75.6	232	2 A61045	homeotic protein 1
38	34	75.6	238	2 T52505	hypothetical prote
39	34	75.6	271	2 A86796	hypothetical prote
40	34	75.6	317	2 T00500	probable elicitor
41	34	75.6	331	1 S48675	carbonate dehydrat
42	34	75.6	355	2 S35345	otxl protein - mou
43	34	75.6	355	2 I56547	homeodomain protei
44	34	75.6	368	2 T01597	hypothetical prote
45	34	75.6	383	2 T38443	hypothetical prote

ALIGNMENTS

RESULT 1

S37923
cell wall protein CWP1 precursor - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YUJ1; protein YKL096w; protein YKL443
C;Species: Saccharomyces cerevisiae
C;Date: 03-May-1994 #sequence revision 03-May-1994 #text change 09-Jul-2004
C;Accession: S37923; S47925; S39090; S4167; PC4042; B57263
R;Cheret, G.; Fukuhara, H.; Bolotin-Fukuhara, M.; Daignan-Fornier, B.; Pallier, C.; Puzos, J.
submitted to the Protein Sequence Database, March 1994
A;Reference number: S37920
A;Accession: S37923
A;Molecule type: DNA
A;Residues: 1-239 <SHI>
A;Cross-references: UNIPROT:P28319; EMBL:Z28096; NID:g486158; PIDN:CAA81934.1; PID:g486158
A;Experimental source: strain S288C
R;Shimoi, H.; Imura, Y.; Obata, T.
submitted to the EMBL Data Library, August 1994
A;Description: Molecular cloning of the gene encoding a yeast cell wall protein, CWP1, w
A;Reference number: S47925
A;Accession: S47925
A;Molecule type: DNA
A;Residues: 1-239 <SHI>
A;Cross-references: EMBL:D37975; NID:g531273; PIDN:BAA07193.1; PID:g1066014
R;Pallier, C.; Valens, M.; Puzos, V.; Fukuhara, H.; Cheret, G.; Sor, F.; Bolotin-Fukuhara, M.
Yeast 9, 1149-1155, 1993
A;Title: DNA sequence analysis of a 17 kb fragment of yeast chromosome XI physically loca
Protein kinases.
A;Reference number: S39084; MUID:94078677; PMID:8256524
A;Accession: S39090
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-239 <PAL>
A;Cross-references: EMBL:X71133; NID:g431205; PIDN:CAA50461.1; PID:g431220
A;Experimental source: strain S288C
R;Forrova, H.; Kolarov, J.; Ghislain, M.; Goffeau, A.
Yeast 8, 419-422, 1992
A;Title: Sequence of the novel essential gene YJU2 and two flanking reading frames locat
A;Reference number: S25354; MUID:92327850; PMID:1626433
A;Accession: S25354
A;Molecule type: DNA
A;Residues: 24-188, 'V', 190-239 <FOR>
A;Cross-references: EMBL:X66245; NID:g4813; PIDN:CAA46969.1; PID:g4814
A;Experimental source: strain S288C
A;Note: the chromosome assignment has been revised
R;Shimoi, H.; Imura, Y.; Obata, T.
J. Biochem. 118, 302-311, 1995
A;Title: Molecular cloning of CWP1: A gene encoding a Saccharomyces cerevisiae cell wall
A;Reference number: JCA167; MUID:96064148; PMID:8543563
A;Accession: JCA167
A;Molecule type: DNA
A;Residues: 1-239 <SHW>
A;Cross-references: DDBJ:D37975; NID:g531273; PIDN:BAA07193.1; PID:g1066014

A:Accession: PC4042
A:Molecule type: protein
A:Residues: 21-139 <SH2>
A:Note: parts of this sequence, including the amino end of the mature protein, were determined by R.Van Der Vaart, J.M.; Caro, L.H.P.; Chapman, J.W.; Klis, F.M.; Verrips, C.T.
J. Bacteriol. 177, 3104-3110, 1995
A:Title: Identification of three mannoproteins in the cell wall of *Saccharomyces cerevisiae*
A:Reference number: A57263; MUID:95286490; PMID:7768807
A:Accession: B57263
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-239 <VAN>
C:Comment: This protein belongs to a new family of cell wall protein. It is a serine-rich cell wall. The carboxy-terminal part plays a crucial role in anchoring to the cell wall
C:Genetics:
A:Gene: SGD:CPW1
A:Cross-references: SGD:S0001579; MIPS:YKL096w
A:Map position: 111
C:Superfamily: serine-rich protein
C:Keywords: cell wall; glycoprotein; transmembrane protein
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-239/Product: cell wall protein CPW1 #status experimental <MAT>
F:222-239/Domain: transmembrane #status predicted <TM>

Query Match 88.9%; Score 40; DB 2; Length 239;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSSSASASSA 10
|||:|||||
Db 154 SSSSASASSA 163

RESULT 2
B86433
hypotheoretical protein T518.20 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: B86433
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: B86433
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-319 <STO>
A:Cross-references: UNIPROT:Q9SA90; GB:AE005172; PIDN:AA25762.1; GSPDB:GN
C:Genetics:
A:Map position: 1

Query Match 86.7%; Score 39; DB 2; Length 319;
Best Local Similarity 81.8%; Pred. No. 19;
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSSSASASSAQ 11
|||:|||||
Db 260 SSSSSSSSAQ 270

RESULT 3
D84629
hypotheoretical protein At2g23820 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C:Accession: D84629

R.; Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, I.; euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J. Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: D84629
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-243 <STO>
A:Cross-references: UNIPROT:O82212; GB:AE002093; NID:g3738315; PIDN:AAC63656.1; GSPDB:GN
C:Genetics:
A:Gene: At2g23820
A:Map position: 2

Query Match 82.2%; Score 37; DB 2; Length 243;
Best Local Similarity 90.0%; Pred. No. 31;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSSSASASSA 10
|||:|||||
Db 70 SSSSASASSA 79

RESULT 4
G86321
F6A14.11 protein - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: G86321
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: G86321
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-269 <STO>
A:Cross-references: UNIPROT:Q9M9U9; GB:AE005172; NID:g6730706; PIDN:AAF27101.1; GSPDB:GN
C:Genetics:
A:Map position: 1

Query Match 82.2%; Score 37; DB 2; Length 269;
Best Local Similarity 90.0%; Pred. No. 34;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSSSASASSA 10
|||:|||||
Db 250 SSSSASASSS 259

RESULT 5
T41394
hypotheoretical serine-rich protein - fission yeast (*Schizosaccharomyces pombe*)
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T41394
R;Wood, V.; Rajandream, M.A.; Barrell, B.G.; Murphy, L.; Harris, D.
submitted to the EMBL Data Library, May 1998
A:Reference number: Z21991
A:Accession: T41394
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-349 <WOO>
A:Cross-references: UNIPROT:O74947; EMBL:AL023704; PIDN:CAA19262.1; GSPDB:GN00068; SPDB:S
A:Experimental source: strain 972h-; cosmid c553

C:Genetics:
A:Gene: SPDB:SPCC553.10
A:Map position: 3

Query Match 82.2%; Score 37; DB 2; Length 349;
Best Local Similarity 90.0%; Pred. No. 43;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSSASASSA 10
|||||:
Db 309 SSSASASSS 318

RESULT 6
S39406
homeotic protein otx1 - human
C:Species: Homo sapiens (man)
C:Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 16-Aug-2004
C:Accession: S39406
R:Simeone, A.; Acampora, D.; Mallamaci, A.; Stornaiuolo, A.; d'Apice, M.R.; Nigro, V.; EMO J. 12, 2735-2747, 1993
A:Title: A vertebrate gene related to orthodenticle contains a homeodomain of the bicoid
A:Reference number: S35345; MUID:93327763; PMID:8101484
A:Accession: S39406
A>Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-354 <SIM>
A:Cross-references: UNIPROT:P32242
C:Genetics:
A:Gene: otx1

C:Superfamily: homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:39-95/Domain: homeobox homology <HOX>
Query Match 82.2%; Score 37; DB 2; Length 354;
Best Local Similarity 90.0%; Pred. No. 44;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSSASASSA 10
|||||:
Db 138 SSSASASSS 147

RESULT 7
I48188
gene Nkx6.1 protein - golden hamster
C:Species: Mesocricetus auratus (golden hamster)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 16-Aug-2004
C:Accession: I48188
R:Rudnick, A.; Ling, T.Y.; Odagiri, H.; Rutter, W.J.; German, M.S.
Proc. Natl. Acad. Sci. U.S.A. 91, 12203-12207, 1994
A:Title: Pancreatic beta cells express a diverse set of homeobox genes.
A:Reference number: I48185; MUID:95083670; PMID:7991607
A:Accession: I48188
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-364 <RES>
A:Cross-references: UNIPROT:Q60554; EMBL:X81409; NID:g587466; PIDN:CAA57166.1; PID:g5874

C:Genetics:
A:Gene: Nkx6.1
C:Superfamily: homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:237-293/Domain: homeobox homology <HOX>

Query Match 82.2%; Score 37; DB 2; Length 364;
Best Local Similarity 90.0%; Pred. No. 45;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSSASASSA 10
|||||:
Db 126 SSSASATSA 135

RESULT 8
S07638

spore coat protein SP96 precursor - slime mold (Dictyostelium discoideum)
C:Species: Dictyostelium discoideum
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 09-Jul-2004
C:Accession: S07638; A60942; B60942
R:Posnaugh, K.L.; Loomis, W.F.
Nucleic Acids Res. 17, 9489, 1989
A:Title: Sequence of the Dictyostelium discoideum spore coat gene SP96.
A:Reference number: S07638; MUID:90067962; PMID:2587278
A:Accession: S07638
A:Molecule type: DNA
A:Residues: 1-600 <POS>
A:Cross-references: UNIPROT:P14328; EMBL:X16491; NID:g7373; PIDN:CAA34508.1; PID:g295736

R:Tasaka, M.; Hasegawa, M.; Ozaki, T.; Iwabuchi, M.; Takeuchi, I.
Cell Differ. Dev. 31, 1-9, 1990
A:Title: Isolation and characterization of spore coat protein (sp96) gene of Dictyostelium
A:Reference number: A60942; MUID:91028801; PMID:1977501
A:Accession: A60942
A:Molecule type: DNA
A:Residues: 1-155, 'T', 157-414, 'C', 416-600 <TAS>
A:Accession: B60942
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-155, 'T', 157-414, 'C', 416-600 <TA2>
C:Genetics:

A:Introns: 22/1
C:Keywords: duplication; glycoprotein
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-600/Product: spore coat protein SP96 #status predicted <MAT>
F:142-185/Region: 15-residue repeats
F:447-533/Region: 9-residue repeats
F:568-600/Region: 32-residue repeats
F:100/Binding site: carbohydrate (Asn) (covalent) #status predicted
Query Match 82.2%; Score 37; DB 2; Length 600;
Best Local Similarity 90.0%; Pred. No. 71;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSSASASSA 10
|||||:
Db 522 SSSASASSS 531

RESULT 9
EDBE11

immediate-early protein IE110 - human herpesvirus 1 (strain 17)
C:Species: human herpesvirus 1
C:Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 09-Jul-2004
C:Accession: A29152
R:Perry, L.J.; Rixon, F.J.; Everett, R.D.; Frame, M.C.; McGeoch, D.J.
J. Gen. Virol. 67, 2365-2380, 1986
A:Title: Characterization of the IE110 gene of herpes simplex virus type 1.
A:Reference number: A29152; MUID:87059760; PMID:3023529
A:Accession: A29152
A:Molecule type: DNA
A:Residues: 1-775 <RES>
A:Cross-references: UNIPROT:P08393; GB:X04614; NID:g59832; PIDN:CAA28285.1; PID:g59833

C:Genetics:
A:Introns: 19/3; 242/1
C:Superfamily: herpesvirus immediate-early protein IE110; RING finger homology
C:Keywords: DNA binding; early protein; transcription regulation; zinc finger
F:112-162/Domain: RING finger homology <RNG>
F:116-156/Region: zinc finger C3HC4 motif

Query Match 82.2%; Score 37; DB 1; Length 775;
Best Local Similarity 90.0%; Pred. No. 91;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSSASASSA 10
|||||:
Db 560 SSSASASSS 569

submitted to the Protein Sequence Database, May 2000

A;Reference number: Z25022
A;Accession: T49804
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1283 <SCH>
A;Cross-references: UNIPROT:Q9GUA3; EMBL:AL356834; GSPDB:GN00116; NCSP:B11B22.60
A;Experimental source: BAC clone B11B22; strain OR74A
C;Genetics:
A;Gene: NCSP:B11B22.60
A;Map position: 6
A;Introns: 856/2

Query Match 82.2%; Score 37; DB 2; Length 1283;
Best Local Similarity 90.0%; Pred. No. 1.5e+02;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSSSASASSA 10
Db 51 SSSSASASSS 60
|||||

RESULT 15

B81067
conserved hypothetical protein NMB1578 [imported] - Neisseria meningitidis (strain MC58)
C;Species: Neisseria meningitidis
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004
C;Accession: B81067
R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
xi, H.; Qin, H.; Vanathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, N.
Science 287, 1809-1815, 2000
A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A;Reference number: A81000; MUID:20175755; PMID:10710307
A;Accession: B81067
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-217 <TET>
A;Cross-references: UNIPROT:Q9JYH9; GB:AE002508; GB:AE002098; NID:g7226820; PIDN:AAF4193
A;Experimental source: serogroup B, strain MC58
C;Genetics:
A;Gene: NMB1578

Query Match 80.0%; Score 36; DB 2; Length 217;
Best Local Similarity 72.7%; Pred. No. 40;
Matches 8; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSSSASASSA 11
Db 32 ASSSASASAAE 42
:|||||

Search completed: July 18, 2005, 13:41:49
Job time : 18.94 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2005, 12:59:21 ; Search time 83.38 Seconds
(without alignments)
67.557 Million cell updates/sec

Title: SEQ2
Perfect score: 45
Sequence: 1 sssasasasaq 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Uniprot 03:.*
1: uniprot_sprot:.*
2: uniprot_trembl:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	42	93.3	361	2 Q6IEP2	Q6iep2 oryza sativ
2	40	88.9	239	1 CWPI_YEAST	P28319 saccharomyc
3	40	88.9	367	1 NK61_HUMAN	P78426 homo sapien
4	39	86.7	169	9 QVVR7	Q9vvr7 drosophila
5	39	86.7	212	2 Q49S2	Q49s2 arabidopsis
6	39	86.7	232	2 Q8S7G5	Q8s7g5 oryza sativ
7	39	86.7	269	2 Q9DAG2	Q9dag2 mus musculu
8	39	86.7	306	2 Q9SA90	Q9sa90 arabidopsis
9	39	86.7	319	2 Q6VNB9	Q6vnb9 mus musculu
10	39	86.7	327	2 Q6VNB9	Q6vnb9 mus musculu
11	39	86.7	361	2 Q6EYI2	Q6eyi2 oryza sativ
12	39	86.7	374	2 Q6EYD1	Q6eyd1 chlamydomon
13	39	86.7	421	2 Q6NR09	Q6nr09 drosophila
14	39	86.7	522	2 Q7PMX8	Q7pmx8 anopheles g
15	39	86.7	851	2 Q6BLS7	Q6bls7 leishmania
16	39	86.7	1405	2 Q7S1R2	Q7s1r2 neurospora
17	37	82.2	131	2 Q6J9R1	Q6j9r1 arabidopsis
18	37	82.2	172	2 Q6VYH6	Q6vyh6 drosophila
19	37	82.2	193	2 Q9D6J0	Q9d6j0 mus musculu
20	37	82.2	228	2 Q69078	Q69078 human herpe
21	37	82.2	243	2 Q9P2R2	Q9p2r2 homo sapien
22	37	82.2	243	2 Q9N2G1	Q9n2g1 pongo pygma
23	37	82.2	243	2 Q9N2G2	Q9n2g2 gorilla gor
24	37	82.2	243	2 Q9N2G3	Q9n2g3 pan troglod
25	37	82.2	243	2 Q82212	Q82212 arabidopsis
26	37	82.2	257	2 Q6DBH4	Q6dbh4 arabidopsis
27	37	82.2	269	2 Q9M9U9	Q9m9u9 arabidopsis
28	37	82.2	280	2 Q869S3	Q869s3 dictyosteli
29	37	82.2	313	2 Q8L5S6	Q8l5s6 oryza sativ
30	37	82.2	349	2 Q74947	Q74947 schizosacch
31	37	82.2	354	1 OTX1_HUMAN	P32242 homo sapien

32	37	82.2	354	2 Q7FN01	Q7fn01 oryza sativ
33	37	82.2	354	2 Q9LDB7	Q9ldb7 oryza sativ
34	37	82.2	364	1 NK61_MESAU	Q60554 mesocricetu
35	37	82.2	365	1 NK61_MOUSE	Q99ma9 mus musculu
36	37	82.2	365	1 NK61_RAT	Q35762 rattus norv
37	37	82.2	408	2 Q75JA3	Q75ja3 dictyosteli
38	37	82.2	421	2 Q84QY7	Q84qy7 oryza sativ
39	37	82.2	431	2 Q9VFP2	Q9vpf2 drosophila
40	37	82.2	520	2 Q6Z9N2	Q6zhn2 oryza sativ
41	37	82.2	532	2 Q6H3Z9	Q6h3z9 oryza sativ
42	37	82.2	600	1 SP96_DICDI	P14328 dictyosteli
43	37	82.2	600	2 Q86B01	Q86b01 dictyosteli
44	37	82.2	619	2 Q960W5	Q960w5 drosophila
45	37	82.2	632	1 APS_HUMAN	O14492 homo sapien

ALIGNMENTS

RESULT 1

Q6IEP2 PRELIMINARY; PRT; 361 AA.
AC Q6IEP2;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE WRKY transcription factor 39.
GN Name=WRKY39;
OS Oryza sativa (indica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaeae; Oryza.
OX NCBI_TaxID=39946;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=15047897; DOI=10.1104/pp.103.034967;
RA Zhang Z.-L., Xie Z., Zou X., Casaretto J., Ho T.-H., Shen Q.J.;
RT "A rice WRKY gene encodes a transcriptional repressor of the
RL gibberellin signaling pathway in aleurone cells.";
RN [2]
RP SEQUENCE FROM N.A.
RA Zhang Z.-L., Xie Z., Zou X., Casaretto J., Ho T.-h.D., Shen Q.J.;
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
CC -!- MISCELLANEOUS: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ third party annotation (TPA) entry.
DR EMBL; BX005042; DAA05104.1; -;
DR GO; GO:0003677; F:DNA binding; IEA.
DR InterPro; IPR003657; WRKY.
DR Pfam; PF03106; WRKY; 1.
DR PROSITE; PS50811; WRKY; 1.
SQ SEQUENCE 361 AA; 38258 MW; 2D87962E300A7BCD CRC64;

Query Match 93.3%; Score 42; DB 2; Length 361;
Best Local Similarity 90.9%; Pred. No. 32;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SSSASASASQAQ 11
Db 251 SSSASASASQAQ 261

RESULT 2

CWPI_YEAST STANDARD; PRT; 239 AA.
ID CWPI_YEAST
AC P28319;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Cell wall protein CWPI precursor.
GN Name=CWPI; OrderedLocustNames=YKL096W; ORFNames=YKL443, YJUL1;
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

```

OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1] SEQUENCE FROM N.A.
RP STRAIN=8288C;
RC MEDLINE=94078677; PubMed=8256524;
RX Pallier C., Valens M., Puzos V., Fukuhara H., Cheret G., Sor F.,
RA Bolotin-Fukuhara M.;
RT "DNA sequence analysis of a 17 kb fragment of yeast chromosome XI
RT physically localizes the MRB1 gene and reveals eight new open reading
RT frames, including a homologue of the KIN1/KIN2 and SNF1 protein
RT kinases.";
RL Yeast 9:1149-1155(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=X2180-1A;
RX MEDLINE=96064148; PubMed=8543563;
RA Shimoi H., Iimura Y., Obata T.;
RT "Molecular cloning of CWPI: a gene encoding a Saccharomyces cerevisiae
RT cell wall protein solubilized with Rarobacter faecitabidus protease
RT I.";
RL J. Biochem. 118:302-311(1995).
RN [3]
RP SEQUENCE OF 24-239 FROM N.A.
RC STRAIN=8288C;
RX MEDLINE=92327850; PubMed=1626433;
RA Fortova H., Kolarov J., Ghislain M., Coffeau A.;
RT "Sequence of the novel essential gene YJ2 and two flanking reading
RT frames located within a 3.2 kb EcoRI fragment from chromosome X of
RT Saccharomyces cerevisiae.";
RL Yeast 8:419-422(1992).
RN [4]
RP SEQUENCE OF 21-30, AND O-GLYCOSYLATION.
RX MEDLINE=95386490; PubMed=7768807;
RA van der Vaart J.M., Caro L.H.P., Chapman J.W., Klis F.M.,
RA Vertips C.T.;
RT "Identification of three mannoproteins in the cell wall of
RT Saccharomyces cerevisiae.";
RL J. Bacteriol. 177:3104-3110(1995).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (potential).
CC -!- PTM: Extensively O-glycosylated.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X71133; CAA50461.1; -
DR EMBL; D37975; BAA07193.1; -
DR EMBL; X66245; CAA46969.1; -
DR EMBL; Z28096; CAA81934.1; -
DR PIR; S37923; S37923.
DR GenOnline; 139852; -
DR COMPUJYEAST-2DPAGE; P28319; -.
DR SGD; S00001579; CWPI.
DR GO; GO:0009277; C:cell wall (sensu Fungi); IDA.
DR GO; GO:0005199; F:structural constituent of cell wall; IDA.
DR GO; GO:0007047; P:cell wall organization and biogenesis; IDA.
DR InterPro; IPR008999; Actin-crosslink.
DR InterPro; IPR000420; Yeast_FIR.
DR Pfam; PF00399; FIR; 1.
DR PROSITE; PS0256; PIR REPEAT 2; 1.
KW Cell wall; Direct protein sequencing; Glycoprotein; GPI-anchor;
LPoprotein; signal.
FT SIGNAL 1 20 Cell wall protein CWPI.
FT CHAIN 21 217 Removed in mature form (Potential).
FT PROPEP 218 239 Ala/Ser-rich.
FT DOMAIN 87 239 GPI-anchor amidated asparagine
FT LIPID 217 217

```

```

FT CONFLICT 189 189 (Potential).
FT SEQUENCE 239 AA; 24268 MW; A85906180D630BAE CRC64;
SQ A -> V (in Ref. 3).

Query Match 88.9%; Score 40; DB 1; Length 239;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSSASASSA 10
Db 154 SSSASASSA 163

RESULT 3
NK61_HUMAN
ID NK61_HUMAN STANDARD; PRT; 367 AA.
AC P78426;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Homeobox protein Nkx-6.1.
GN Name=NKX6-1; Synonyms=NKX6A;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancratic islets;
RX MEDLINE=97237060; PubMed=9119408; DOI=10.1006/geno.1996.4568;
RA Inoue H., Rudnick A., German M.S., Veille R., Donis-Keller H.,
RA Permutt M.A.;
RT "Isolation, characterization, and chromosomal mapping of the human
RT NKx6.1 gene (NKX6A), a new pancreatic islet homeobox gene.";
RL Genomics 40:367-370(1997).
CC -!- FUNCTION: May be important for control of islet development and/or
CC regulation of insulin biosynthesis.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- TISSUE SPECIFICITY: Pancratic beta cells.
CC -!- SIMILARITY: Contains 1 homeobox domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U66799; AAD11962.1; -
DR EMBL; U66797; AAD11962.1; JOINED.
DR EMBL; U66798; AAD11962.1; JOINED.
DR HSSP; PL3297; 1IG7.
DR TRNSPAC; T04268; -.
DR Genew; HGNC:7839; NKX6-1.
DR MIM; 602563; -.
DR GO; GO:0009887; P:organogenesis; TAS.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR009057; Homeodomain-like.
DR InterPro; IPR000047; HTH_lambrepresr.
DR Pfam; PF00046; Homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.
DR PRINTS; PR00031; HTHREPRESSR.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX 1; 1.
DR PROSITE; PS00071; HOMEBOX 2; 1.
KW Developmental protein; DNA-binding; Homeobox; Nuclear protein.
FT DOMAIN 49 61 Poly-Ser.
FT DOMAIN 119 135 Poly-Ser.
FT DOMAIN 136 150 Poly-Ala.
FT DOMAIN 168 173 Poly-Pro.
FT DNA_BIND 236 295 Homeobox.

```


FT DOMAIN 351 355 Poly-Gly.
 SQ SEQUENCE 367 AA; 37848 MW; EIB2F06E3E046483 CRC64;
 Query Match 88.9%; Score 40; DB 1; Length 367;
 Best Local Similarity 100.0%; Pred. No. 70;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSSASASSA 10
 Db 127 SSSASASSA 136
 |||||
 |||||

RESULT 4
 Q9VRR7 PRELIMINARY; PRT; 169 AA.
 AC Q9VRR7
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE CGL3290-PA.
 GN ORFNames=CGL3290;
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=20196006; PubMed=107311132; DOI=10.1126/science.287.5461.2185;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
 RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Binkley J., Brodeur G.M.,
 RA Burtis K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodagef, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195 (2000).
 RN [2]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=22426065; PubMed=12537568;
 RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
 RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
 RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,

RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
 RA Svirkas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
 RA Weinstock G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
 RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila
 RT melanogaster euchromatic genome sequence.";
 RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
 RN [3]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=22426070; PubMed=12537573;
 RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirkas R.,
 RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
 RA Ashburner M., Celniker S.E.;
 RT "The transposable elements of the Drosophila melanogaster euchromatin:
 RT a genomics perspective.";
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084 (2002).
 RN [4]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=22426069; PubMed=12537572;
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Milburn G.H., Prochuk S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a
 RT systematic review.";
 RL Genome Biol. 3:RESEARCH0083-RESEARCH0083 (2002).
 RN [5]
 RN SEQUENCE FROM N.A.
 RP SEQUENCE FROM N.A.
 RG FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RN SEQUENCE FROM N.A.
 RG FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB003564; AAF50724.1; ..
 SQ SEQUENCE 169 AA; 18990 MW; D63E2979A8C4D255 CRC64;
 Query Match 86.7%; Score 39; DB 2; Length 169;
 Best Local Similarity 81.8%; Pred. No. 45;
 Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSSASASSA 11
 Db 21 SSSASASSA 31
 |||||
 |||||

RESULT 5
 Q949S2 PRELIMINARY; PRT; 212 AA.
 ID Q949S2
 AC Q949S2
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE Hypothetical protein Atig30750.
 GN Names=Atig30750;
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopses.
 OX NCBI_TaxID=3702;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
 RA Goldsmith A.D., Lee J.M., Quach H.D., Toriumi M., Yu G., Bowser L.,
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
 RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
 RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.

```

RN  SEQUENCE FROM N.A.
RP  Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,
RA  Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,
RA  Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,
RA  Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA  Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA  Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
RA  Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
RA  Theologis A.; (2002) to the EMBL/GenBank/DBJ databases.
RL  EMBL: AY059922; AAK93599.1; -.
DR  EMBL: AY113970; AAM45018.1; -.
KW  Hypothetical protein.
SQ  SEQUENCE 212 AA; 22213 MW; 851CD2920F35D414 CRC64;

Query Match      86.7%; Score 39; DB 2; Length 212;
Best Local Similarity 81.8%; Pred. NO. 57;
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY  1 SSSSSASSSAQ 11
Db  153 SSSSSSSSSSAQ 163

RESULT 6
Q8S7G5 PRELIMINARY; PRT; 232 AA.
AC  Q8S7G5;
DT  01-JUN-2002 (TrEMBLrel. 21, Created)
DT  01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE  Putative transmembrane protein.
GN  Name=OSNBB0048D20.8;
OS  Oryza sativa (Rice).
OC  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC  Ehrhartoideae; Oryzeae; Oryza.
OX  NCBI_TaxID=4530;
RN  SEQUENCE FROM N.A.
RP  Buell C.R., Yuan Q., Ouyang S., Liu J., Moffat K.S., Hill J.N.,
RA  Gansberger K., Brenner M., Burgess S., Hance M., Shvartsbeyn M.,
RA  Tsitrin T., Riggs F., Hsiao J., Ziemann V., Blunt S., Pai G.,
RA  VanAken S.E., Utterback T.R., Feldblyum T.V., Kalb E., Quackenbush J.,
RA  Salzberg S.L., White O., Fraser C.M.;
RL  Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
RN  SEQUENCE FROM N.A.
RP  Buell R.;
RL  Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR  EMBL: AC084766; AL82519.1; -.
DR  Gramene; Q8S7G5; -.
DR  GO: GO:0016021; C:integral to membrane; IEA.
DR  GO: GO:0008320; F:protein carrier activity; IEA.
DR  GO: GO:0006886; P:intracellular protein transport; IEA.
DR  InterPro: IPR000348; EMP24_sp25f_p24.
DR  InterPro: IPR009038; GOLD.
DR  Pfam: PF01105; EMP24_GP25L; 1.
DR  PROSITE: PS08666; GOLD; 1.
KW  Transmembrane.
SQ  SEQUENCE 232 AA; 25829 MW; 7215F83F243F1187 CRC64;

Query Match      86.7%; Score 39; DB 2; Length 232;
Best Local Similarity 81.8%; Pred. NO. 63;
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY  1 SSSSSASSSAQ 11
Db  67 SSSSSSSSSSAQ 77

RESULT 7

```

```

Q9DAG2 PRELIMINARY; PRT; 269 AA.
AC  Q9DAG2;
DT  01-JUN-2001 (TrEMBLrel. 17, Created)
DT  01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT  01-WAR-2004 (TrEMBLrel. 26, Last annotation update)
DE  Mus musculus adult male testis cDNA, RIKEN full-length enriched
DE  library, clone:1700011H05 product:hypothetical Serine-rich region
DE  containing protein, full insert sequence.
GN  Name=Spats1;
OS  Mus musculus (Mouse).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX  NCBI_TaxID=10090;
RN  SEQUENCE FROM N.A.
RP  STRAIN=C57BL/6J; TISSUE=Testis;
RC  MEDLINE=9927253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RX  Carninci P., Hayashizaki Y.;
RA  "High-efficiency full-length cDNA cloning.";
RL  Meth. Enzymol. 303:19-44(1999).
RN  SEQUENCE FROM N.A.
RP  STRAIN=C57BL/6J; TISSUE=Testis;
RC  MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RX  RIKEN FANTOM Consortium;
RA  "Functional annotation of a full-length mouse cDNA collection.";
RL  Nature 409:685-690(2001).
RN  SEQUENCE FROM N.A.
RP  STRAIN=C57BL/6J; TISSUE=Testis;
RC  MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RX  RIKEN FANTOM Consortium;
RA  "Functional annotation of a full-length mouse cDNA collection.";
RL  Nature 409:685-690(2001).
RN  SEQUENCE FROM N.A.
RP  STRAIN=C57BL/6J; TISSUE=Testis;
RC  MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RX  Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA  Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT  "Normalization and subtraction of cap-trapper-selected cDNAs to
RT  prepare full-length cDNA libraries for rapid discovery of new genes.";
RL  Genome Res. 10:1617-1630(2000).
RN  SEQUENCE FROM N.A.
RP  STRAIN=C57BL/6J; TISSUE=Testis;
RC  MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RX  Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA  Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA  Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA  Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA  Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA  Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
RA  Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT  "RIKEN integrated sequence analysis (RISA) system-384-format
RT  sequencing pipeline with 384 multicapillary sequencer.";
RL  Genome Res. 10:1757-1771(2000).
RN  SEQUENCE FROM N.A.
RP  STRAIN=C57BL/6J; TISSUE=Testis;
RC  STRAIN=C57BL/6J; TISSUE=Testis;
RX  Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA  Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA  Hangaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoaka T., Hori F.,
RA  Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA  Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
RA  Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA  Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai K., Sakai K.,
RA  Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA  Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA  Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA  Muramatsu M., Hayashizaki Y.;

```



```

ID Q6VNB9 PRELIMINARY; PRT; 327 AA.
AC Q6VNB9;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Dv12-DEP domain interacting protein.
GN Name=Ddip;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Embryo;
RA Ng S.S., Han L., Zhang X., He X., Chang Z.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
SQ EMBL; AY336501; AAQ8273.1; -.
SEQUENCE 327 AA; 36420 MW; 4433D94631091274 CRC64;

Query Match 86.7%; Score 39; DB 2; Length 327;
Best Local Similarity 81.8%; Pred. No. 90;
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSSSASASSAQ 11
Db 68 SSSSSSSSSAQ 78

RESULT 11
Q6ETY2
ID Q6ETY2 PRELIMINARY; PRT; 361 AA.
AC Q6ETY2;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Putative WRKY transcription factor.
GN Name=P000608.18-1;
OS Eryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF004683; BAD2788.1; -.
DR GO; GO:0003677; F:DNA binding; IEA.
DR InterPro; IPR003657; WRKY.
DR Pfam; PF03106; WRKY; 1.
DR PROSITE; PS50811; WRKY; 1.
SQ SEQUENCE 361 AA; 38242 MW; 7D9869D99167CD16 CRC64;

Query Match 86.7%; Score 39; DB 2; Length 361;
Best Local Similarity 81.8%; Pred. No. 1e+02;
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSSSASASSAQ 11
Db 251 SSSAASASSAQ 261

RESULT 12
Q66YD1
ID Q66YD1 PRELIMINARY; PRT; 374 AA.
AC Q66YD1;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Chloroplast DnaJ-like protein 2.
GN Name=CDJ2;
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;

```

```

OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RA Liu C., Schroda M.;
RT "J-domain protein CDJ2 and HSP70B are a plastidic chaperone pair that
interacts with vesicle inducing proteins in plastids 1 (VIPP1).";
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY696657; AAU06581.1; -.
DR InterPro; IPR001623; DnaJ_N.
DR Pfam; PF0226; DnaJ; 1.
DR PRINTS; PR00625; DNAJPROTEIN.
DR SMART; SM00271; DnaJ; 1.
DR PROSITE; PS00636; DnaJ_1; 1.
DR PROSITE; PS50076; DnaJ_2; 1.
SQ SEQUENCE 374 AA; 38400 MW; 0C9EB3E4B8642544 CRC64;

Query Match 86.7%; Score 39; DB 2; Length 374;
Best Local Similarity 81.8%; Pred. No. 1e+02;
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSSSASASSAQ 11
Db 322 SSSSASSSSSQ 332

RESULT 13
Q6NR09
ID Q6NR09 PRELIMINARY; PRT; 421 AA.
AC Q6NR09;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE RE08107p.
GN Name=CG13290;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
Celniker S.;
RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BT010275; AAQ23593.1; -.
DR HSSP; P35235; IAYD.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR InterPro; IPR000980; SH2.
DR Pfam; PF00017; SH2; 1.
DR PRINTS; PR00401; SH2DOMAIN.
DR ProDom; PD000093; SH2; 1.
DR SMART; SM00252; SH2; 1.
DR PROSITE; PS50001; SH2; 1.
SQ SEQUENCE 421 AA; 46098 MW; 67729D2C824DD3F0 CRC64;

Query Match 86.7%; Score 39; DB 2; Length 421;
Best Local Similarity 81.8%; Pred. No. 1.2e+02;
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSSSASASSAQ 11
Db 21 SSSSASSSSSQ 31

RESULT 14
Q7PMX8

```

```

ID Q7PMX8 PRELIMINARY; PRT; 522 AA.
AC Q7PMX8;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE ENSANGP00000019885 (Fragment).
GN Name=ENSANGG00000017396;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAB01008966; EAA13079.2; -.
FT NON_TER 1
SQ SEQUENCE 522 AA; 58307 MW; 8B4268DB806FDE0A CRC64;

Query Match 86.7%; Score 39; DB 2; Length 522;
Best Local Similarity 81.8%; Pred. No. 1.5e+02;
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSSSSASASQAQ 11
Db 58 SSSSSSSSSQAQ 68
||||:|||||

RESULT 15
Q9BLS7
ID Q9BLS7 PRELIMINARY; PRT; 851 AA.
AC Q9BLS7;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein L6071.04 (Fragment).
GN Name=L6071.04;
OS Leishmania major.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Friedlin;
RX MEDLINE=98146435; PubMed=9477341;
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
RA Smith D.F.;
RT "A physical map of the Leishmania major Friedlin genome.";
RL Genome Res. 8:135-145(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Friedlin;
RA Zimmermann W., Ivens A.C., Quail M., Rajandream M.A., Barrell B.G.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL583933; CAC32263.1; -.
KW Hypothetical protein.
FT NON_TER 851
SQ SEQUENCE 851 AA; 92071 MW; 4CE78FF1416D742B CRC64;

Query Match 86.7%; Score 39; DB 2; Length 851;
Best Local Similarity 81.8%; Pred. No. 2.5e+02;
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSSSSASASQAQ 11
Db 210 SSSSSSSSSQAQ 220
||||:|||||

```

Search completed: July 18, 2005, 13:33:19
Job time : 86.38 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2005, 12:59:21 ; Search time 92.62 Seconds
(without alignments)
45.934 Million cell updates/sec

Title: SEQ2

Perfect score: 45

Sequence: 1 sssasasaaq 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq_16Dec04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	100.0	11	3 AAY71087	Synthetic
2	40	88.9	239	8 ADS43814	Bacterial
3	40	88.9	367	5 ABB10101	Human hom
4	40	88.9	367	5 ABG70892	Human NKX
5	40	88.9	367	8 ADQ09851	Human NKX
6	39	86.7	69	4 AAU32281	Novel hum
7	39	86.7	189	4 ABB67983	Drosophil
8	39	86.7	251	8 ADP98907	C. albica
9	39	86.7	361	8 ADI42062	Plant tra
10	39	86.7	361	8 ADO02607	Thalecres
11	38	84.4	283	5 AAU19446	Hansenula
12	37	82.2	109	4 AAU18125	Novel hum
13	37	82.2	109	4 AAU18675	Renal and
14	37	82.2	109	4 ABB10317	Human cdn
15	37	82.2	109	4 AAU87616	Novel cen
16	37	82.2	109	4 AAU18486	Human end
17	37	82.2	109	4 ADU19684	Protein e
18	37	82.2	109	5 ABJ05752	Novel hum
19	37	82.2	109	5 ABP66904	Human pol
20	37	82.2	109	6 ABU97290	Human pol
21	37	82.2	109	8 ADI54931	Novel hum
22	37	82.2	170	7 ABO67531	Klebsiell
23	37	82.2	215	5 AAU17826	Herpes si
24	37	82.2	215	7 ADG74974	Human her
25	37	82.2	224	4 ABB65466	Drosophil

26	37	82.2	255	6 ABU70608	Human adi
27	37	82.2	257	4 AAM80270	Human pro
28	37	82.2	354	4 AAM79286	Human pro
29	37	82.2	361	8 ADR86098	Aspergill
30	37	82.2	365	8 ADQ09852	Mouse NK-
31	37	82.2	431	4 ABB69274	Drosophil
32	37	82.2	431	8 ADS96458	Drosophil
33	37	82.2	632	6 ADA10932	Human cdn
34	37	82.2	632	7 ADE55442	Human pro
35	37	82.2	632	7 ADD45189	Human pro
36	37	82.2	775	7 ADJ91994	Human her
37	37	82.2	825	6 ABP71596	HSV-2 ICP
38	37	82.2	825	7 ADG75089	Human her
39	37	82.2	826	5 AAU17827	Herpes si
40	37	82.2	826	7 ADG74975	Human her
41	37	82.2	980	5 ABB91667	Herbicida
42	37	82.2	1059	4 ABB65792	Drosophil
43	37	82.2	1077	4 ABB61539	Drosophil
44	37	82.2	1318	4 AAB26874	HSV-2 imm
45	37	82.2	1318	7 ADG75125	Human her

ALIGNMENTS

RESULT 1

AAAY71087

ID AAY71087 standard; peptide; 11 AA.

XX AC AAY71087;

XX AC AAY71087;

DT 21-SEP-2000 (first entry)

XX XX

DE Synthetic linker peptide #2 encoded by MV03JA oligonucleotide linker.

XX KW

XX KW

KW Iiams; HC-V; heavy chain variable domain; antigen binding protein;

KW linker; conformational flexibility; multivalent binding protein; bi-head;

KW human chorionic gonadotropin; hCG; immunoassay; agglutination assay;

KW purification.

XX OS Synthetic.

XX OS Synthetic.

FH Key

FT Peptide

FT /label= Peptide linker 2

FT /note= "Flanked by one residue from N- and C-terminii of

FT HCV fragment"

XX WO200024884-A2.

PN 04-MAY-2000.

XX 22-OCT-1999; 99WO-EP008323.

XX 27-OCT-1998; 98WO-EP006991.

PR 22-APR-1999; 99EP-00303118.

XX (UNIL) UNILEVER PLC.

PA (UNIL) UNILEVER NV.

PA (HIND-) HINDUSTAN LEVER LTD.

XX Frenken LGJ, Howell S, Van Der Vaart JM;

XX WPI; 2000-350728/30.

XX N-PSDB; AAD00659.

XX Use of a linker whose amino acid sequence confers restricted

PT conformational flexibility to generate multivalent and multispecific

PT antigen binding proteins.

XX Example 1.1d; Page 20; 50pp; English.

XX The present sequence is the synthetic linker peptide #2, encoded by the

CC

CC oligonucleotide linker fragment, MV03UA. It consists of the last residue
 CC of the N-terminal HC-V fragment (S) and the first residue of the C-
 CC terminal HC-V fragment (O), intersected by the connecting linker peptide.
 CC It is used for the construction of *Saccharomyces cerevisiae* episomal
 CC expression plasmid, pUR531, encoding anti-HCG-anti-RR6 bispecific
 CC biheads, containing the linker peptide. The peptide linker confers
 CC restricted conformational flexibility for linking binding units in a
 CC multivalent binding protein. The linker is used to generate multivalent
 CC or multispecific antigen binding proteins for immunoassays, agglutination
 CC assays or for purification
 XX
 SQ Sequence 11 AA;

Query Match 100.0%; Score 45; DB 3; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.19; Mismatches 0; Gaps 0;
 Matches 11; Conservative 0; Indels 0; Gaps 0;

QY 1 SSSSSASSAAQ 11
 |||||
 Db 1 SSSSSASSAAQ 11

RESULT 2
 ADS43814
 ID ADS43814 standard; protein; 239 AA.

XX ADS43814;

DT 02-DEC-2004 (first entry)

DE Bacterial polypeptide #22244.

XX Recombinant DNA construct; transformed plant; improved plant property;
 KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
 KW pathogen tolerance; pest tolerance; plant disease resistance;
 KW cell cycle pathway modification; plant growth regulator;
 KW homologous recombination; seed oil yield; protein yield; carbohydrate;
 KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
 KW bacterial polypeptide.

XX Bacteria.

XX US2003233675-A1.

PN 18-DEC-2003.

XX 20-FEB-2003; 2003US-00369493.

XX 21-FEB-2002; 2002US-0360039P.

XX (CAOY/) CAO Y.
 PA (HINK/) HINKLE G J.
 PA (SLAT/) SLATER S C.
 PA (CHEN/) CHEN X.
 PA (GOLD/) GOLDMAN B S.

XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
 PI WPI; 2004-061375/06.

XX New recombinant DNA construct comprising a promoter positioned to provide
 PT for expression of a polynucleotide encoding a polypeptide from a
 PT microbial source, useful for producing plants with improved properties.
 XX Claim 1; SEQ ID NO 22244; 122pp; English.

XX The invention relates to a recombinant DNA construct comprising a
 CC promoter functional in a plant cell, where the promoter is positioned to
 CC provide for expression of a polynucleotide encoding a polypeptide from a
 CC microbial source. The invention also relates to a transformed plant
 CC comprising the recombinant DNA construct and a method of producing a
 CC transformed plant having an improved property. The plant is a crop plant
 CC such as maize or soybean. The method of producing a transformed plant

CC having an improved property comprises transforming a plant with the
 CC recombinant DNA construct and growing the transformed plant, where the
 CC polynucleotide or polypeptide is useful for improving plant properties.
 CC The recombinant DNA construct is useful for producing plants with
 CC improved plant properties, e.g. improved cold, heat or drought tolerance,
 CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
 CC increased resistance to plant disease, better growth rate by modification
 CC of the cell cycle pathway with plant growth regulators, increased rate of
 CC homologous recombination, modified seed oil or protein yield and/or
 CC content, improved yield by modification of carbohydrate, nitrogen or
 CC phosphorus use and/or uptake, by modification of photosynthesis or by
 CC providing improved plant growth and development under at least one stress
 CC condition, improved lignin production or improved galactomannan
 CC production. This sequence represents a bacterial polypeptide used in the
 CC scope of the invention. Note: The sequence data for this patent did not
 CC form part of the printed specification but was obtained in electronic
 CC format from USPTO at seqdata.uspto.gov/sequence.html.
 XX
 SQ Sequence 239 AA;

Query Match 88.9%; Score 40; DB 8; Length 239;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSSSASASSA 10
 |||||

Db 154 SSSSASASSA 163

RESULT 3
 ABB10101

ID ABB10101 standard; protein; 367 AA.

XX ABB10101;

DT 01-JUL-2002 (first entry)

DE Human homeobox protein Nkx6.1 amino acid sequence.

XX Human; homeobox protein Nkx6.1; homeobox protein Nkx6.2; ventral neuron;
 KW amyotrophic lateral sclerosis; spinal muscular atrophy;
 KW neurodegenerative disease; transcription factor.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 224

FT /note= "histidine residue that is not encoded in either
 of the exon 1 or exon 2 containing DNA fragments (see
 ABL56817 and ABL56818)"

XX WO200218545-A1.

XX 07-MAR-2002.

XX 31-AUG-2001; 2001WO-US027256.

XX 01-SEP-2000; 2000US-00654462.

XX (UYCO) UNIV COLUMBIA NEW YORK.

XX Jessell TM, Briscoe J, Ericson J, Rubenstein JLR, Sander M;
 DR N-PSDB; ABL56817, ABL56818, ABL56819.

XX WPI; 2002-329764/36.

XX Converting stem cell into ventral neuron useful for treating neural
 PT degeneration in a subject, by introducing nucleic acid or polypeptide
 PT expressing homeodomain transcription factor Nkx6.1 or Nkx6.2 protein.

XX Disclosure; Fig 6; 108pp; English.

XX The invention relates to a method for converting a stem cell (SC) into a

CC ventral neuron, comprising introducing a nucleic acid or polypeptide
CC expressing homeodomain transcription factor Nkx6.1 or Nkx6.2 protein into
CC a stem cell. Methods of the invention are useful for converting a stem
CC cell into a ventral neuron, and for diagnosing a motor neuron
CC degenerative disease such as amyotrophic lateral sclerosis or spinal
CC muscular atrophy in a subject. Methods of the invention are also useful
CC for treating neuronal degeneration in a subject. The method comprises
CC implanting a neural stem cell in diseased neural tissue under conditions
CC so that the stem cell is converted into motor neuron after implantation.
CC The current sequence represents the human homeobox protein Nkx6.1 amino
CC acid sequence
XX
SQ Sequence 367 AA;

Query Match 88.9%; Score 40; DB 5; Length 367;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSSSSASASSA 10
Db 127 SSSSSASASSA 136
|||||

RESULT 4
ABG70892
ID ABG70892 standard; protein; 367 AA.
XX
AC ABG70892;
XX
DT 04-DEC-2002 (first entry)
XX
DE Human NKX-6.1 protein.
XX
KW Human; NKX-6.1; diabetes; depression; obesity; hyperglycaemia;
KW Parkinson's disease; neural defect disorder; chromosome 4q21.2-4q22;
KW reduced serotonin production; developmental defect.
XX
OS Homo sapiens.
XX
PN US6436667-B1.
XX
PD 20-AUG-2002.
XX
PF 20-JAN-1998; 98US-00009816.
XX
PR 25-JUL-1997; 97US-00900510.
XX
PA (REGC) UNIV CALIFORNIA.
PA (UNIW) UNIV WASHINGTON.
XX
PI German MS, Permutt MA, Inoue H;
XX
DR WPI; 2002-711490/77.
DR N-PSDB; ABS54691.
XX
XX
PT Novel human Nkx6.1 polynucleotide encoding Nkx6.1 polypeptide useful for
PT treating diseases associated with Nkx6.1 polypeptide activity, e.g., Type
PT 1 and Type 2 diabetes, Parkinson's disease, depression and obesity.
XX
PS Claim 6; Fig 3; 34pp; English.
XX

The invention relates to an isolated polynucleotide or its complement
CC (1), comprising a polynucleotide sequence: (a) that is at least 95%
CC identical to a polynucleotide of human NKX-6.1 where (i) encodes a
CC polypeptide that promotes development of pancreatic beta cells; or (b)
CC encoding human NKX-6.1 polypeptide sequence. Also included are NKX-6.1
CC recombinant expression vectors, host cells and probes. The host cells are
CC useful for producing a human NKX-6.1 polypeptide, NKX-6.1 nucleic acid
CC is useful for producing NKX-6.1 polypeptide or its fragments by
CC recombinant techniques, identifying and isolating polynucleotide and
CC polypeptide sequence having homology to human NKX-6.1 polypeptide, for
CC identifying human NKX-6.1 polypeptide binding compounds and for
CC diagnosing, preventing and treating (e.g. by gene therapy) diseases

CC associated with human NKX-6.1 polypeptide biological activity, such as
CC disorders associated with reduced levels of insulin or the ability to
CC utilise insulin (e.g. hyperglycaemia, diabetes (e.g. Type 1 and Type 2
CC diabetes), Parkinson's disease, disorders associated with reduced
CC serotonin production (e.g. depression and obesity), and disorders
CC associated with neural defects (e.g. defects in motor neurons, serotonin-
CC producing neurons, dopamine neurons, and developmental defects in the
CC forebrain, midbrain, hindbrain and spinal cord). The gene for human NKX-
CC 6.1 is located on chromosome 4q21.2-4q22. The present sequence represents
XX human NKX-6.1
SQ Sequence 367 AA;

Query Match 88.9%; Score 40; DB 5; Length 367;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSSSSASASSA 10
Db 127 SSSSSASASSA 136
|||||

RESULT 5
ADQ09851
ID ADQ09851 standard; protein; 367 AA.
XX
AC ADQ09851;
XX
DT 23-SEP-2004 (first entry)
XX
DE Human NK-2 class homeobox protein, NKx6.1.
XX
KW Human; islet cell differentiation transcription factor;
KW insulin-dependent diabetes; insulin; somatic cell;
KW insulin-producing cell; islet cell; NeuroD; ngn3; neurogenin 3; Pax6;
KW paired-box transcription factor 6; Pax4;
KW paired-box transcription factor 4; Nkx2.2; NK-2 class homeobox protein;
KW NKx6.1; Isl-1; Islet factor 1; Pdx-1;
KW pancreatic and duodenal homeobox protein; BCT; betacellulin.
XX
OS Homo sapiens.
XX
PN US2004132679-A1.
XX
PD 08-JUL-2004.
XX
PF 03-SEP-2003; 2003US-00654102.
XX
PR 03-SEP-2002; 2002US-0407743P.
XX
PA (BAYU) BAYLOR COLLEGE MEDICINE.
XX
PI Chan L, Kojima H;
XX
DR WPI; 2004-517032/49.
DR N-PSDB; ADQ09870, ADQ09871, ADQ09872, ADQ09944, ADQ09945, ADQ09946.
XX
PT Use of an islet cell differentiation transcription factor polypeptide or
PT its homologue or analog for treating a mammal for insulin-dependent
PT diabetes, increasing an insulin level in a somatic cell, or generating an
PT insulin-producing cell.
XX
PS Claim 109; SEQ ID NO 89; 190pp; English.
XX

The invention relates to the use of an islet cell differentiation
CC transcription factor polypeptide or its homologue or analogue for
CC treating a mammal for insulin-dependent diabetes, increasing an insulin
CC level in a somatic cell, or generating an insulin-producing cell. Also
CC included are a method of treating a mammal for insulin-dependent
CC diabetes, a method of increasing an insulin level in a somatic cell, a
CC method of generating an insulin-producing cell, a therapeutic composition
CC comprising an isolated islet cell differentiation transcription factor
CC polypeptide (and/or an isolated nucleic acid expressing the polypeptide),

CC an insulin-producing cell comprising a vector (where the vector comprises
 CC a nucleic acid sequence encoding an islet cell differentiation
 CC transcription factor), an insulin-producing cell (generated by a method
 CC comprising obtaining a somatic cell and transfecting the cell with a
 CC vector comprising a nucleic acid sequence encoding an islet cell
 CC differentiation transcription factor, where in the transfecting step the
 CC cell produces insulin), a method of generating at least one pancreatic
 CC islet, and a composition (comprising: NeuroD or ngn3 polypeptide or a
 CC polynucleotide expressing a NeuroD or ngn3 polypeptide and betacellulin
 CC polypeptide or a polynucleotide expressing a betacellulin polypeptide).
 CC The islet cell differentiation transcription factor polypeptide is
 CC NeuroD, ngn3 (neurogenin 3), Pax6 (paired-box transcription factor 6),
 CC Pax4 (paired-box transcription factor 4), Nkx2.2 (NK-2 class homeobox
 CC protein), Nkx6.1, Isl-1 (islet factor 1), Pdx-1 (pancreatic and duodenal
 CC homeobox protein), BCT (betacellulin) or their combinations. The islet
 CC cell differentiation transcription factor polypeptide or its homologue or
 CC analogue is useful for treating a mammal for insulin-dependent diabetes,
 CC increasing an insulin level in a somatic cell, or generating an insulin-
 CC producing cell. The present sequence is an islet cell differentiation
 CC transcription factor polypeptide as detailed above.
 XX
 XX
 SQ Sequence 367 AA;

Query Match 88.9%; Score 40; DB 8; Length 367;
 Best Local Similarity 100.0%; Pred. No. 66;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSSSASASSA 10
 DB 127 SSSSASASSA 136
 |||||

RESULT 6
 AAU32281
 ID AAU32281 standard; protein; 69 AA.
 XX
 AC AAU32281;
 DT
 DT 18-DEC-2001 (first entry)
 XX
 DE Novel human secreted protein #2772.

XX Human; vaccination; gene therapy; nutritional supplement;
 KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
 KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
 XX
 XX Homo sapiens.

XX WO200179449-A2.
 XX
 XX 25-OCT-2001.

PF 16-APR-2001; 2001WO-US008656.

XX 18-APR-2000; 2000US-00552929.

PR 26-JAN-2001; 2001US-00770160.

XX (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-611725/70.

XX Nucleic acids encoding a range of human polypeptides, useful in genetic
 PT vaccination, testing and therapy.

XX Claim 20; Page 575; 765pp; English.

XX The invention relates to novel human secreted polypeptides. The
 CC polypeptides and antibodies to the polypeptides are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of polypeptide. The polypeptides are also useful for
 CC identifying agents (agonists and antagonists) that bind to them. Cells

CC expressing the proteins are useful for identifying a therapeutic agent
 CC for use in treatment of a pathology related to aberrant expression or
 CC physiological interactions of the polypeptide. Vectors comprising the
 CC nucleic acids encoding the polypeptides and cells genetically engineered
 CC to express them are also useful for producing the proteins. The proteins
 CC are useful in genetic vaccination, testing and therapy, and can be used
 CC as nutritional supplements. They may be used to increase stem cell
 CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
 CC and/or nerve tissue growth or regeneration; immune suppression and/or
 CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
 CC secreted proteins of the invention
 XX
 SQ Sequence 69 AA;

Query Match 86.7%; Score 39; DB 4; Length 69;
 Best Local Similarity 81.8%; Pred. No. 15;
 Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSSSASASSA 11
 DB 52 SSSSSSSSAQ 62
 |||||

RESULT 7
 ABB67983
 ID ABB67983 standard; protein; 169 AA.

XX ABB67983;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 30741.

KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

XX (PEXE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR N-PSDB; ABL12086.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.

XX Disclosure; SEQ ID NO 30741; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 169 AA;

Query Match 86.7%; Score 39; DB 4; Length 169;
 Best Local Similarity 81.8%; Pred. No. 42;
 Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSSSSASSSAQ 11
 Db 21 SSSSSSSSSQ 31
 |||||:|:|

RESULT 8
 ADP98907
 ID ADP98907 standard; protein; 251 AA.
 XX
 AC ADP98907;
 DT 23-SEP-2004 (first entry)
 XX
 XX C. albicans specific gene, orf6.1768, protein sequence.
 XX
 XX Diploid fungal cell; allele; gene disruption cassette;
 KW promoter replacement fragment; antifungal; fungicide; gene therapy;
 KW infection; Candida albicans.
 XX
 OS Candida albicans.
 XX
 XX WO2004056965-A2.
 PN
 XX 08-JUL-2004.
 PD
 XX 19-DEC-2003; 2003WO-US040618.
 PF
 XX 19-DEC-2002; 2002US-0434832P.
 PR
 XX (ELIT-) ELITRA PHARM INC.
 PA (ELIT-) ELITRA CANADA LTD.
 XX
 XX Roemer T, Jiang B, Boone C, Bussey H;
 PI WPI; 2004-500296/47;
 XX N-PSDB; ADP98597.
 DR
 XX Constructing a strain of diploid fungal cells in which both alleles of a
 PT gene are modified comprises modifying the alleles of a gene in the fungal
 PT cells by recombination using a gene disruption cassette and a promoter
 PT replacement fragment.
 XX
 PS Claim 44; SEQ ID NO 7082; 163pp; English.
 XX
 XX The invention relates to a novel method for constructing a strain of
 CC diploid fungal cells in which both alleles of a gene are modified. The
 CC method comprises modifying the alleles of a gene in diploid fungal cells
 CC by recombination using a gene disruption cassette and a promoter
 CC replacement fragment. The invention further comprises: assembling a
 CC collection of diploid fungal cells each of which comprises modified
 CC alleles of a different gene; a strain of diploid fungal cells comprising
 CC modified alleles of a gene, where the first allele of the gene is
 CC inactivated by a gene disruption cassette comprising a nucleotide
 CC sequence encoding an expressible selectable marker; and the expression of
 CC the second allele of the gene is regulated by a heterologous promoter.
 CC That is operably linked to the coding region of the second allele of the
 CC gene, and where the gene encodes the polypeptide mentioned above; a
 CC collection of diploid fungal strains comprising the diploid strains cited
 CC above, where substantially all the different genes that encode the above
 CC amino acid sequences are modified and are present in different diploid
 CC strains in the collection; a nucleic acid molecule microarray comprising
 CC nucleic acid molecules, where each nucleic acid molecule comprises a
 CC nucleotide sequence that is hybridizable to a target nucleotide sequence
 CC comprising any of the 310 nucleotide sequences listed in the
 CC specification (ADP98516-ADP98825); identifying a gene that is essential
 CC to the survival or growth of a fungus, that contributes to the virulence
 CC and/or pathogenicity of a fungus, or that contributes to the resistance
 CC of a diploid fungus to an antifungal agent; identifying an antifungal

agent that inhibits the growth of a diploid fungus, or a therapeutic
 levels of proteins or gene transcripts with the inhibition of growth or
 proliferation of a diploid fungal cell; a purified or isolated nucleic
 acid molecule comprising a nucleotide sequence encoding a gene product
 required for proliferation of Candida albicans, where the gene product
 consists of any of the above-mentioned amino acid sequences; a vector
 comprising a promoter operably linked to the nucleic acid molecule cited
 above; a host cell containing the vector; a purified or isolated
 polypeptide comprising any of the 61 amino acid sequences given in the
 specification (ADP96718-ADP96778); a fusion protein comprising a fragment
 of a first polypeptide fused to a second polypeptide, the fragment
 consisting of at least 6 consecutive residues of any of ADP98826-ADP99135
 ; producing a polypeptide; identifying a compound which modulates the
 activity of a gene product encoded by a nucleic acid comprising any of
 ADP98516-ADP98825; eliciting an immune response in an animal; a strain of
 Candida albicans, where a first allele of a gene comprising any of
 ADP98516-ADP98825 is inactive and a second allele of the gene is under
 the control of a heterologous promoter; identifying a compound or binding
 partner that binds to the polypeptide comprising any of ADP98826-
 ADP99135, or its fragment; identifying a compound having the ability to
 inhibit growth or proliferation of Candida albicans; inhibiting growth or
 proliferation of Candida albicans cells; manufacturing an antimycotic
 compound; treating an infection of a subject by Candida albicans;
 preventing or containing contamination of an object by Candida albicans,
 or for preventing or inhibiting formation on a surface of a biofilm
 comprising Candida albicans; a pharmaceutical composition comprising a
 therapeutic amount of an agent which reduces the activity or level of a
 gene product encoded by a nucleic acid comprising any of ADP98516-
 ADP98825 in a pharmaceutical carrier; an antibody preparation which binds
 the polypeptide; methods for evaluating a compound against a target gene
 product encoded by any of ADP98516-ADP98825; identifying an antimycotic
 compound; a computer or a computer readable medium that comprises at
 least one of the nucleotide sequences mentioned in the specification or
 at least one amino acid sequence selected from ADP98826-ADP99135; a
 method assisted by a computer for identifying a putatively essential gene
 of a fungus; and a protein array comprising proteins, where at least one
 protein comprises an amino acid sequence or a portion of an amino acid
 sequence selected from ADP98516-ADP98825. The novel methods and
 compositions have fungicide activity. The compositions may be used in
 gene therapy. The composition and methods are useful for drug screening
 purposes or for diagnosing, preventing or treating infections associated
 with Candida albicans. These may also be used for constructing strains
 useful for identification and validation of gene products as effective
 targets for therapeutic intervention, for identifying and validating gene
 products as effective targets for therapeutic intervention, and for
 collecting identified essential genes. This sequence represents the
 protein of a Candida albicans fungal specific gene of the invention.
 CC NOTE: This sequence was downloaded from an electronic sequence listing
 CC provided on the WIPO website.
 XX
 SQ Sequence 251 AA;

Query Match 86.7%; Score 39; DB 8; Length 251;
 Best Local Similarity 81.8%; Pred. No. 65;
 Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSSSSASSSAQ 11
 Db 133 SSSSSSSSSSAQ 143
 |||||:|:|

RESULT 9
 ADI42062
 ID ADI42062 standard; protein; 361 AA.
 XX
 AC ADI42062;
 XX
 DT 22-APR-2004 (first entry)
 XX
 XX Plant transcription factor #244.
 DE
 XX transgenic; plant; enhanced tolerance to abiotic stress;

XX Zhang J, Fromm ME, Heard JE, Riechmann JL, Adam LJ, Broun PE;
 PI Pineda O, Reuber TL, Keddle JS, Yu G, Jiang C, Samaha RS;
 PI Filgrim ML, Creelman RA, Dubell AN, Ratcliffe O, Kumimoto R;
 PI Sherman BK;
 XX WPI; 2004-225755/21.
 XX
 XX New transgenic plant, useful in developing phenotypes with altered or
 PT improved characteristics or traits.
 XX
 XX Claim 1; SEQ ID NO 1021; 213pp; English.
 XX
 XX The invention relates to a transgenic plant comprising a recombinant
 CC polynucleotide having a polynucleotide sequence or its complementary
 CC sequence comprising a sequence encoding a polypeptide, that initiates
 CC transcription (i.e. a transcription factor) from Arabidopsis, Soybean,
 CC Rice, Rape or Corn, comprising any of the sequences appearing as ADO01588
 CC -ADO03527 or ADO03530-ADO03559. Also included are using a transgenic
 CC plant to grow a progeny plant, an expression cassette comprising a
 CC constitutive, inducible or tissue-specific promoter and a recombinant
 CC polynucleotide described above), a host cell comprising the expression
 CC cassette, producing a modified plant having a modified trait, identifying
 CC a factor that is modulated by or interacts with a polypeptide encoded by
 CC the polynucleotide sequence and identifying at least one downstream
 CC polynucleotide sequence that is subject to a regulatory effect of any of
 CC the polypeptides encoded by the polynucleotide described above. The
 CC transgenic plant is useful for producing a plant that has an altered
 CC trait e.g. an enhanced tolerance to abiotic stress (increased tolerance
 CC to chilling, germination in cold conditions, freezing tolerance, tolerance
 CC to heat, tolerance to drought, tolerance to osmotic stress, tolerance to
 CC salt, tolerance to phosphate limitation, tolerance to potassium
 CC limitation, decreased sensitivity to nitrogen limitation), altered
 CC hormone sensitivity, reduced sensitivity to abscisic acid, an altered
 CC response to ethylene, disease resistance, altered susceptibility to
 CC Botrytis, altered susceptibility to Fusarium, altered susceptibility to
 CC Erysiphe, altered susceptibility to Pseudomonas syringae, altered
 CC susceptibility to Sclerotinia, altered sugar sensing, improved seed
 CC germination and seedling vigor, early flowering, late flowering, extended
 CC period of flowering, an inflorescence architectural change, a change in
 CC stem bifurcations, a lack of a shoot meristem, reduced meristem cell
 CC differentiation, altered phyllotaxy, altered branching pattern, reduced
 CC apical dominance, reduced trichome density, ectopic trichome development,
 CC altered trichome development, altered stem morphology, increased root
 CC growth, increased root hairs, altered seed development, altered cell
 CC proliferation/cell differentiation, premature senescence, delayed
 CC senescence, lethality, increased necrosis, an increase in seedling or
 CC plant size, decreased plant size, a change in leaf morphology, increased
 CC altered leaf development, increased leaf size and mass, glossy leaves,
 CC leaf cell expansion, change in seed morphology, altered seed coloration,
 CC increased seed size, decreased seed size, altered seed shape, change in
 CC leaf biochemistry, increased leaf wax, an alteration in leaf prenyl lipid
 CC content, increased leaf insoluble sugars, decreased leaf insoluble
 CC sugars, increased leaf anthocyanins, an alteration of leaf fatty acid
 CC content, an alteration of leaf glucosinolate content, change in seed
 CC biochemistry, an increase in seed oil content, decrease in seed oil
 CC content, increase in seed fatty acid content, decrease in seed fatty acid
 CC content, increase in seed protein content, decrease in seed protein
 CC content, alteration in seed prenyl lipid content, increase in seed
 CC sterols, upregulation of genes involved in secondary metabolism, increase
 CC in root anthocyanins, increase in plant anthocyanins, and alteration in
 CC light response or shade avoidance. The present sequence represents an
 CC orthologue of a thalecress transcription factor isolated from Rice.
 XX
 XX Sequence 361 AA;

Query Match 86.7%; Score 39; DB 8; Length 361;
 Best Local Similarity 81.8%; Pred. No. 97;
 Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SSSASASSAQ 11
 |||||:
 Db 251 SSSASASSAQ 261

RESULT 11

AAE19446

XX AAE19446 standard; protein; 283 AA.

AC AAE19446;

XX 29-AUG-2002 (revised)

DT 31-MAY-2002 (first entry)

XX

DE Hansenula polymorpha GPI cell wall anchor protein, Hptip1p.

XX

KW Cell wall anchor protein; HpsED1; HpPIR2; HpgAS1; Hptip1; HpcWPI; GPI;
 KW glycosyl phosphatidyl inositol; surface expression system; industry;
 KW medical; food; chemical; biochemical.

OS Pichia angusta.

XX WO200212509-A1.

PN 14-FEB-2002.

XX

PF 27-JUL-2000; 2000WO-KR000819.

XX

PR 26-JUL-2000; 2000KR-00042939.

XX (KORE-) KOREA RES INST BIOSCIENCE & BIOTECHNOLOG.

PA Choi E, Sohn J, Kim S;

XX WPI; 2002-227157/28.

DR N-PSDB; AAD30795.

XX

PT Novel cell wall protein genes HpsED1, HpPIR2, HpgAS1, Hptip1, HpcWPI

PT derived from yeast Hansenula polymorpha, which encode the cell wall

PT proteins useful for construction of surface expression systems.

XX Claim 11; Page 114-116; 121pp; English.

XX The present invention relates to a cell wall protein gene derived from

CC Hansenula polymorpha, such as HpsED1, HpPIR2, HpgAS1, Hptip1, HpcWPI. The

CC surface expression systems have numerous applications including

CC immobilisation of biocatalyst and large scale production of protein such

CC as enzymes, antigens, antibodies, etc. The surface expression systems

CC comprising Hptip1p, HpsED1p, HpgAS1p, HpcWpi1p, are useful for exporting

CC glucose oxidase to cell surface. The surface expression systems have

CC applications in medical industry, food industry and chemical and

CC biochemical industry. The surface proteins derived from H. polymorpha are

CC highly effective in construction of surface expression systems and

CC development of biocatalyst application systems. The present sequence is

CC Hansenula polymorpha GPI (glycosyl phosphatidyl inositol) cell wall

CC anchor protein, Hptip1p. (Updated on 29-AUG-2003 to standardise OS field)

XX Sequence 283 AA;

Query Match 84.4%; Score 38; DB 5; Length 283;

Best Local Similarity 81.8%; Pred. No. 1.1e+02;

Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SSSASASSAQ 11

|||||:
 Db 141 SSSASASSAQ 151

XX

RESULT 12

AAU18125

XX AAU18125 standard; protein; 109 AA.

AC AAU18125;

XX 21-NOV-2001 (first entry)

DT

XX

DE Novel human uterine motility-association polypeptide #32.
XX Human; uterine motility-association disorder; uterus; pregnancy; labour;
KW menstrual cycle; gene therapy.
XX
OS Homo sapiens.
PN WO20015201-A1.
XX 02-AUG-2001.
PD
XX
PF 17-JAN-2001; 2001WO-US001317.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 01-SEP-2000; 2000US-0228287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239355P.
PR 20-OCT-2000; 2000US-023937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 01-DEC-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 06-DEC-2000; 2000US-0256719P.
PR 08-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.

```

PR 08-DEC-2000; 2000US-0251950P.
PR 11-DEC-2000; 2000US-0254037P.
PR 03-JAN-2001; 2001US-0259678P.
XX
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-488777/53.
XX N-PSDB; AAS28967.
XX
XX New nucleic acid molecules encoding 49 human secreted proteins for
XX diagnosing, preventing, treating or ameliorating medical conditions and
XX used as food additives or preservatives.
XX
XX Claim 11; SEQ ID NO 101; 524pp; English.
XX
XX The present invention relates to the isolation of novel human uterine
XX motility-association polypeptides, and cDNA (AAS28936-AAS28994) and
XX genomic sequences encoding for these polypeptides. The sequences of the
XX invention are useful in the diagnosis, treatment, prevention and/or
XX prognosis of diseases associated with uterine motility such as pregnancy
XX and labour, and menstrual disorders. The polynucleotide sequences of the
XX invention are also useful in gene therapy. AAU18094-AAU18152 represent
XX novel human uterine motility-association polypeptides. Note: The sequence
XX data for this patent did not form part of the printed specification, but
XX was obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 109 AA;
SQ
Query Match 82.2%; Score 37; DB 4; Length 109;
Best Local Similarity 90.0%; Pred. No. 56;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 SSSSSASASSA 10
Db 96 SSSSSASASSA 105
||||:|||||
||||:|||||

RESULT 13
AAU18675
ID AAU18675 standard; protein; 109 AA.
XX
XX AC AAU18675;
XX
XX 21-NOV-2001 (first entry)
XX
XX Renal and cardiovascular-associated protein, Seq ID 114.
XX
XX Human; antiinflammatory; neuroprotective; immunomodulator; vulnery;
XX cardiovascular; cytostatic; nephrotropic; antianaemic; nephritis;
XX immunosuppressive; kidney disorder; renal failure; hypertension;
XX cardiovascular disorder; myocardial infarction; blood disorder; anaemia;
XX blood coagulation disorder; electrolyte imbalance disorder; cancer;
XX hyponatraemia; hyperkalaemia; neoplastic disorder; nephroma;
XX autoimmune disease; inflammatory disease; reproductive system disorder;
XX endocrine disorder; neural activity; neurological disorder;
XX wound healing; respiratory disorder.
XX
XX Homo sapiens.
XX
XX WO200155328-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US0001359.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180628P.
XX 24-FEB-2000; 2000US-0184664P.
XX 02-MAR-2000; 2000US-0186350P.
XX 16-MAR-2000; 2000US-0189874P.
XX
17-MAR-2000; 2000US-0190076P.
18-APR-2000; 2000US-0198123P.
19-MAY-2000; 2000US-0205515P.
20-JUN-2000; 2000US-0209467P.
21-JUN-2000; 2000US-0214886P.
22-JUN-2000; 2000US-0215135P.
23-JUL-2000; 2000US-0216647P.
24-JUL-2000; 2000US-0216880P.
25-JUL-2000; 2000US-0217487P.
26-JUL-2000; 2000US-0217496P.
27-JUL-2000; 2000US-0218290P.
28-JUL-2000; 2000US-0220963P.
29-JUL-2000; 2000US-0220964P.
30-AUG-2000; 2000US-0224518P.
31-AUG-2000; 2000US-0224519P.
14-AUG-2000; 2000US-0225213P.
14-AUG-2000; 2000US-0225214P.
14-AUG-2000; 2000US-0225266P.
14-AUG-2000; 2000US-0225267P.
14-AUG-2000; 2000US-0225268P.
14-AUG-2000; 2000US-0225270P.
14-AUG-2000; 2000US-0225447P.
14-AUG-2000; 2000US-0225757P.
14-AUG-2000; 2000US-0225758P.
14-AUG-2000; 2000US-0225759P.
18-AUG-2000; 2000US-0226279P.
22-AUG-2000; 2000US-0226681P.
22-AUG-2000; 2000US-0226686P.
23-AUG-2000; 2000US-0227182P.
23-AUG-2000; 2000US-0227009P.
30-AUG-2000; 2000US-0228924P.
01-SEP-2000; 2000US-0229287P.
01-SEP-2000; 2000US-0229343P.
01-SEP-2000; 2000US-0229344P.
05-SEP-2000; 2000US-0229345P.
05-SEP-2000; 2000US-0229509P.
05-SEP-2000; 2000US-0229513P.
06-SEP-2000; 2000US-0230437P.
08-SEP-2000; 2000US-0230438P.
08-SEP-2000; 2000US-0231242P.
08-SEP-2000; 2000US-0231243P.
08-SEP-2000; 2000US-0231244P.
08-SEP-2000; 2000US-0231413P.
08-SEP-2000; 2000US-0231414P.
08-SEP-2000; 2000US-0232080P.
12-SEP-2000; 2000US-0232081P.
14-SEP-2000; 2000US-0231968P.
14-SEP-2000; 2000US-0232397P.
14-SEP-2000; 2000US-0232398P.
14-SEP-2000; 2000US-0232399P.
14-SEP-2000; 2000US-0232400P.
14-SEP-2000; 2000US-0232401P.
14-SEP-2000; 2000US-0233063P.
14-SEP-2000; 2000US-0233064P.
21-SEP-2000; 2000US-0233065P.
21-SEP-2000; 2000US-0234223P.
21-SEP-2000; 2000US-0234274P.
25-SEP-2000; 2000US-0234997P.
25-SEP-2000; 2000US-0234998P.
26-SEP-2000; 2000US-0235484P.
27-SEP-2000; 2000US-0235834P.
27-SEP-2000; 2000US-0235836P.
29-SEP-2000; 2000US-0236327P.
29-SEP-2000; 2000US-0236367P.
29-SEP-2000; 2000US-0236368P.
29-SEP-2000; 2000US-0236369P.
29-SEP-2000; 2000US-0236370P.
02-OCT-2000; 2000US-0236802P.
02-OCT-2000; 2000US-0237037P.
02-OCT-2000; 2000US-0237038P.
02-OCT-2000; 2000US-0237039P.
13-OCT-2000; 2000US-0237040P.
13-OCT-2000; 2000US-0239935P.

```

```

13-OCT-2000; 2000US-0239937P.
20-OCT-2000; 2000US-0240960P.
20-OCT-2000; 2000US-0241221P.
20-OCT-2000; 2000US-0241785P.
20-OCT-2000; 2000US-0241786P.
20-OCT-2000; 2000US-0241787P.
20-OCT-2000; 2000US-0241808P.
20-OCT-2000; 2000US-0241809P.
20-OCT-2000; 2000US-0241826P.
01-NOV-2000; 2000US-0244617P.
08-NOV-2000; 2000US-0246474P.
08-NOV-2000; 2000US-0246475P.
08-NOV-2000; 2000US-0246476P.
08-NOV-2000; 2000US-0246477P.
08-NOV-2000; 2000US-0246478P.
08-NOV-2000; 2000US-0246523P.
08-NOV-2000; 2000US-0246524P.
08-NOV-2000; 2000US-0246525P.
08-NOV-2000; 2000US-0246526P.
08-NOV-2000; 2000US-0246527P.
08-NOV-2000; 2000US-0246528P.
08-NOV-2000; 2000US-0246532P.
08-NOV-2000; 2000US-0246609P.
08-NOV-2000; 2000US-0246610P.
08-NOV-2000; 2000US-0246611P.
08-NOV-2000; 2000US-0246613P.
17-NOV-2000; 2000US-0249207P.
17-NOV-2000; 2000US-0249208P.
17-NOV-2000; 2000US-0249209P.
17-NOV-2000; 2000US-0249210P.
17-NOV-2000; 2000US-0249211P.
17-NOV-2000; 2000US-0249212P.
17-NOV-2000; 2000US-0249213P.
17-NOV-2000; 2000US-0249214P.
17-NOV-2000; 2000US-0249215P.
17-NOV-2000; 2000US-0249216P.
17-NOV-2000; 2000US-0249217P.
17-NOV-2000; 2000US-0249218P.
17-NOV-2000; 2000US-0249244P.
17-NOV-2000; 2000US-0249245P.
17-NOV-2000; 2000US-0249264P.
17-NOV-2000; 2000US-0249265P.
17-NOV-2000; 2000US-0249297P.
17-NOV-2000; 2000US-0249299P.
17-NOV-2000; 2000US-0249300P.
01-DEC-2000; 2000US-0250160P.
01-DEC-2000; 2000US-0250391P.
05-DEC-2000; 2000US-0251030P.
05-DEC-2000; 2000US-0251398P.
05-DEC-2000; 2000US-0256719P.
06-DEC-2000; 2000US-0251479P.
08-DEC-2000; 2000US-0251856P.
08-DEC-2000; 2000US-0251868P.
08-DEC-2000; 2000US-0251869P.
08-DEC-2000; 2000US-0251989P.
08-DEC-2000; 2000US-0251990P.
11-DEC-2000; 2000US-0254097P.
05-JAN-2001; 2001US-0259678P.
(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI; 2001-488787/53.
N-PSDB; AAS30196.
New polynucleotides and polypeptides, useful for diagnosing, treating,
preventing or prognosing e.g. kidney, cardiovascular, blood, electrolyte
imbalance or neoplastic disorders, autoimmune diseases, cancers.
Claim 1; SEQ ID NO 114; 506pp; English.
The invention relates to novel nucleic acids and polypeptides useful for
```

```

CC diagnosing, treating, preventing and/or prognosing disorders related to
CC these polypeptides. The polynucleotides are especially useful in the
CC diagnosis, prognosis, prevention and/or treatment of diseases which
CC include kidney disorders (e.g. renal failure or nephritis),
CC cardiovascular disorders (e.g. hypertension or myocardial infarction),
CC blood disorders (e.g. anaemia or blood coagulation disorders),
CC electrolyte imbalance disorders (e.g. hyponatraemia or hyperkalaemia),
CC neoplastic disorders (e.g. nephroma or renal cell cancer), autoimmune
CC diseases, cancers, inflammatory diseases, reproductive system disorders,
CC endocrine disorders, neural activity and neurological disorders, wound
CC healing and respiratory disorders. AAU18644-AAU18715 represent the novel
CC human renal and cardiovascular-associated amino acid sequences of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at: ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 109 AA;
Query Match 82.2%; Score 37; DB 4; Length 109;
Best Local Similarity 90.0%; Pred. No. 56;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 SSSSSASSA 10
Db 96 SSSSSASSA 105
||||:|||||
RESULT 14
ABBI0317
ID ABBI0317 standard; protein; 109 AA.
XX
AC ABBI0317;
XX
DT 10-JAN-2002 (first entry)
XX
DE Human cDNA SEQ ID NO: 625.
XX
KW Human; gene therapy; neural disorder; immune system disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
KW proliferative disorder; inflammation.
XX
OS Homo sapiens.
XX
PN WO200154474-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001349.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
```


PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241321P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.

PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-476161/51.
DR N-PSDB; ABA06539.
XX
XX Isolated nucleic acid molecule encoding an inflammation-associated
PT polypeptide is used in preventing, treating or ameliorating a medical
PT condition.
XX
XX Claim 11; SEQ ID NO 625; 859pp + Sequence Listing; English.
XX
XX The present invention provides human cDNAs, proteins and related genomic
CC DNAs. These can be used in the treatment of neural, immune system,
CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC renal and proliferative disorders and inflammation. The present sequence
CC is a protein of the invention
XX
SQ Sequence 109 AA;

Query Match 82.2%; Score 37; DB 4; Length 109;
Best Local Similarity 90.0%; Pred. No. 56;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY . . . 1 SSSSSASSA 10
SY . . . |||:||||
Db . . . 96 SSSSSASSA 105

RESULT 15
AAU87616


```

PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-581633/65.
XX N-PSDB; ABK43946.
XX
XX New isolated nucleic acid encoding a protein for diagnosing, preventing,
XX treating or ameliorating medical conditions and used as food additives or
XX preservatives.
XX
XX Claim 9; SEQ ID NO 1134; 837pp; English.
XX
XX The invention describes an isolated nucleic acid molecule (I) encoding a
XX novel central nervous system protein. (I) and polypeptides (III) encoded
XX by (I), are used to treat a medical conditions and in diagnosis of a
XX pathological condition. Disorders which are diagnosed or treated include
XX autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
XX disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
XX e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
XX angiogenesis, nervous system disorders e.g. Alzheimer's disease and
XX amyotrophic lateral sclerosis, infections caused by bacteria, viruses
XX e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders
XX e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,
XX adenocarcinomas and irritable bowel syndrome, reproductive system
XX disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes
XX and pituitary dwarfism, cancers and disorders at the cellular level e.g.
XX leukaemia, disorders involving neovascularisation e.g. malignancies,
XX respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
XX acute kidney failure and blood related disorders e.g. myocardial
XX infarction. The polypeptides can also be used to aid wound healing and
XX epithelial cell proliferation, to prevent skin aging due to sunburn, and
XX maintain organs before transplantation, for supporting cell culture of
XX primary tissues, to regenerate tissues and in chemotaxis. The
XX polypeptides can also be used as a food additive or preservative to
XX increase or decrease storage capabilities, fat content, lipid, protein,
XX
XX Query Match 82.2%; Score 37; DB 4; Length 109;
XX Best Local Similarity 90.0%; Pred. No. 56;
XX Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0
XX
XX QY 1 SSSSSASASSA 10
XX DB 96 SSSSSASASSA 105
XX
XX Search completed: July 18, 2005, 13:40:27
XX Job time : 95.62 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2005; 13:09:46 ; Search time 13.86 Seconds
(without alignments)
62.478 Million cell updates/sec

Title: SEQ3

Perfect score: 49

Sequence: 1 sgspgpggq 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR 79:*

2: Pirl:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	46	93.9	306	2 T21938	hypothetical prote
2	44	89.8	1096	2 A26879	pullulanase (EC 3.
3	43	87.8	266	2 T22706	hypothetical prote
4	43	87.8	299	2 T22705	hypothetical prote
5	43	87.8	299	2 T24833	hypothetical prote
6	43	87.8	302	2 T32872	hypothetical prote
7	43	87.8	304	2 T22482	hypothetical prote
8	43	87.8	316	2 S08169	collagen col-13 pr
9	43	87.8	316	2 T22827	hypothetical glyci
10	43	87.8	356	2 T20993	hypothetical prote
11	43	87.8	463	2 B70893	hypothetical prote
12	42	85.7	208	2 T15245	collagen, cuticula
13	42	85.7	286	2 S34665	hypothetical prote
14	42	85.7	294	2 T22639	hypothetical prote
15	42	85.7	300	2 T24482	hypothetical prote
16	42	85.7	310	2 T22641	hypothetical prote
17	42	85.7	323	2 T19142	hypothetical prote
18	42	85.7	339	2 T22607	hypothetical prote
19	42	85.7	1142	2 JX0369	collagen alpha 1(X
20	42	85.7	1453	2 S21626	collagen alpha 1(I
21	42	85.7	1691	1 CGH06B	collagen alpha 6(I
22	42	85.7	1744	2 S40391	collagen alpha 1(I
23	42	85.7	1758	2 T29350	hypothetical prote
24	42	85.7	1759	2 T29351	collagen alpha 2(I
25	42	85.7	1775	2 A31893	collagen alpha 1(I
26	41	83.7	152	2 T24064	hypothetical prote
27	41	83.7	220	2 AD2990	conserved hypothet
28	41	83.7	220	2 E98293	hypothetical prote
29	41	83.7	302	2 T21257	hypothetical prote

30 41 83.7 307 2 T18846 hypothetical prote
31 41 83.7 325 2 T16324 hypothetical prote
32 41 83.7 371 2 E88633 protein F5683.1 I
33 41 83.7 636 2 S41067 collagen alpha 1(I
34 41 83.7 671 1 CGRT1S collagen alpha 1(I
35 41 83.7 886 2 I50694 collagen alpha 1(I
36 41 83.7 1042 1 CGCH16 collagen alpha 1(I
37 41 83.7 1464 1 CGHUIS collagen alpha 1(I
38 41 83.7 1464 2 S59856 collagen alpha 1(I
39 41 83.7 1486 1 B40333 collagen alpha 1(I
40 41 83.7 1492 2 A40333 collagen alpha 1(I
41 41 83.7 1669 1 CGMS4B collagen alpha 1(I
42 41 83.7 1747 2 A54121 collagen alpha-4 c
43 40 81.6 239 1 LNM5MA mannose-binding le
44 40 81.6 291 2 T20083 hypothetical prote
45 40 81.6 291 2 T26576 hypothetical prote

ALIGNMENTS

RESULT 1

T21938

hypothetical protein F38A3.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C:Accession: T21938

R:Swinsburne, J.

submitted to the EMBL Data Library, June 1995

A:Reference number: Z19490

A:Accession: T21938

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-306 <WIL>

A:Cross-references: UNIPROT:Q20135; EMBL:Z49938; PIDN:CAA90187.1; GSPDB:GN00020; CESP:F38A3

A:Experimental source: clone F38A3

C:Genetics:

A:Gene: CESP:F38A3.1

A:Map position: 2

A:Introns: 26/2; 54/3

Query Match 93.9%; Score 46; DB 2; Length 306;
Best Local Similarity 88.9%; Pred. No. 5;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGSPGSPGQ 9

||:|||||

Db 136 SGNFGSPGQ 144

RESULT 2

A26879

pullulanase (EC 3.2.1.41) precursor - Klebsiella pneumoniae

N:Alternate names: pullulan 6-glucanohydrolase; pullulanase

C:Species: Klebsiella pneumoniae

C:Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 09-Jul-2004

C:Accession: A26879; S02472

R:Katsuragi, N.; Takizawa, N.; Murooka, Y.

J. Bacteriol. 169, 2301-2306, 1987

A:Title: Entire nucleotide sequence of the pullulanase gene of Klebsiella aerogenes W70.

A:Reference number: A26879; MUID:87194626; PMID:3155373

A:Note: K. aerogenes

A:Accession: A26879

A:Molecule type: DNA

A:Residues: 1-1096 <KAT>

A:Cross-references: UNIPROT:P07811

A:Experimental source: strain W70

R:Charalambous, B.M.; Keen, J.N.; McPherson, M.J.

EMBO J. 7, 2903-2909, 1988

A:Title: Collagen-like sequences stabilize homotrimers of a bacterial hydrolase.

A:Reference number: S02472; MUID:89030658; PMID:2846288

A:Accession: S02472

A:Status: not compared with conceptual translation

A:Molecule type: DNA
A:Residues: 1-74 <CHA>
A:Note: part of this sequence was confirmed by protein sequencing
C:Genetics:

A:Gene: pula

C:Superfamily: pullulanase type debranching enzyme

F;1-109/Domain: signal sequence #status predicted <SIG>

F;20-1996/Product: alpha-dextrin endo-1,6-alpha-glucosidase #status predicted <MAT>

Query Match 89.8%; Score 44; DB 2; Length 1096;

Best Local Similarity 100.0%; Pred. No. 35;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGSPGSPG 8

Db 29 SGSPGSPG 36

RESULT 3

T22706

hypothetical protein F55C10.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C:Accession: T22706

R:Dobson, R.

submitted to the EMBL Data Library, June 1996

A:Reference number: Z19603

A:Accession: T22706

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-266 <WIL>

A:Cross-references: UNIPROT:Q21184; EMBL:Z74036; PIDN:CAA98487.1; GSPDB:GN00023; CBSP:F5

A:Experimental source: clone F55C10

C:Genetics:

A:Gene: CBSP:F55C10.3

A:Map position: 5

A:Introns: 21/3

Query Match

Best Local Similarity 87.8%; Score 43; DB 2; Length 266;

Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGSPGSPG 9

Db 135 NGNPGSPG 143

RESULT 4

T22705

hypothetical protein F55C10.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C:Accession: T22705

R:Dobson, R.

submitted to the EMBL Data Library, June 1996

A:Reference number: Z19603

A:Accession: T22705

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-299 <WIL>

A:Cross-references: UNIPROT:Q20805; EMBL:Z74036; PIDN:CAA98486.1; GSPDB:GN00023; CBSP:F5

A:Experimental source: clone F55C10

C:Genetics:

A:Gene: CBSP:F55C10.2

A:Map position: 5

A:Introns: 54/3

Query Match

Best Local Similarity 87.8%; Score 43; DB 2; Length 299;

Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGSPGSPG 9

Db 135 NGNPGSPG 143

Db 168 NGNPGSPG 176

RESULT 5

T24833

hypothetical protein T11F9.9 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C:Accession: T24833

R:Lennard, N.

submitted to the EMBL Data Library, June 1996

A:Reference number: Z19941

A:Accession: T24833

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-299 <WIL>

A:Cross-references: UNIPROT:Q22393; EMBL:Z74042; PIDN:CAA98525.1; GSPDB:GN00023; CBSP:T11

A:Experimental source: clone T11F9

C:Genetics:

A:Gene: CBSP:T11F9.9

A:Map position: 5

A:Introns: 54/3

Query Match

Best Local Similarity 87.8%; Score 43; DB 2; Length 299;

Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGSPGSPG 9

Db 168 NGNPGSPG 176

RESULT 6

T32872

hypothetical protein W05G11.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004

C:Accession: T32872

R:Davidson, S.; Langston, Y.; O'Neal, D.

submitted to the EMBL Data Library, December 1997

A:Description: The sequence of C. elegans cosmid W05G11.

A:Reference number: Z21238

A:Accession: T32872

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-302 <DAV>

A:Cross-references: UNIPROT:O44904; EMBL:AF040660; PIDN:AACT71147.1; GSPDB:GN00021; CBSP:V

A:Experimental source: strain Bristol N2; clone W05G11

C:Genetics:

A:Gene: CBSP:W05G11.3

A:Map position: 3

Query Match

Best Local Similarity 87.8%; Score 43; DB 2; Length 302;

Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGSPGSPG 9

Db 162 AGSPGAPGQ 170

RESULT 7

T22482

hypothetical protein F52B11.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C:Accession: T22482

R:Matthews, L.

submitted to the EMBL Data Library, November 1996

A:Reference number: Z19569

A:Accession: T22482

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

```
A;Residues: 1-304 <WIL>
A;Cross-references: UNIPROT:Q9XUE9; EMBL:Z82268; PIDN:CAB05195.1; GSPDB:GN00022; CESP:F5
A;Experimental source: clone F52B11
C;Genetics:
A;Gene: CESP:F52B11.4
A;Map position: 4
A;Introns: 27/3

Query Match      87.8%; Score 43; DB 2; Length 304;
Best Local Similarity 77.8%; Pred. No. 14;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGSPGSPGQ 9
Db 194 SGAPGPGQ 202

RESULT 8
S08169
collagen col-12 precursor - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 09-Jul-2004
C;Accession: S08169; T20994
R;Park, Y.S.; Kramer, J.M.
J. Mol. Biol. 211, 395-406, 1990
A;Title: Tandemly duplicated Caenorhabditis elegans collagen genes differ in their modes
A;Reference number: S08169; MUID:90172409; PMID:1689778
A;Accession: S08169
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-316 <PAR>
A;Cross-references: UNIPROT:P20630; GB:X51622; NID:96679; PIDN:CAA35954.1; PID:96680
R;Berks, M.
submitted to the EMBL Data Library, June 1996
A;Reference number: Z19356
A;Accession: T20994
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-316 <WIL>
A;Cross-references: EMBL:Z73972; PIDN:CAA98257.1; GSPDB:GN00023; CESP:F15H10.1
A;Experimental source: clone F15H10
C;Genetics:
A;Gene: col-12
A;Map position: 5
A;Introns: 71/3
F;1-36/Domain: signal sequence #status predicted <SIG>
F;37-316/Product: collagen col-12 #status predicted <MAT>

Query Match      87.8%; Score 43; DB 2; Length 316;
Best Local Similarity 77.8%; Pred. No. 15;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGSPGSPGQ 9
Db 190 SGAPGPGQ 198

RESULT 9
S08170
collagen col-13 precursor - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 09-Jul-2004
C;Accession: S08170; T20995
R;Park, Y.S.; Kramer, J.M.
J. Mol. Biol. 211, 395-406, 1990
A;Title: Tandemly duplicated Caenorhabditis elegans collagen genes differ in their modes
A;Reference number: S08169; MUID:90172409; PMID:1689778
A;Accession: S08170
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-316 <PAR>
A;Cross-references: UNIPROT:P20631; GB:X51623; NID:96681; PIDN:CAA35955.1; PID:96682
R;Berks, M.
```

```
submitted to the EMBL Data Library, June 1996
A;Reference number: Z19356
A;Accession: T20995
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-316 <WIL>
A;Cross-references: EMBL:Z73972; PIDN:CAA98258.1; GSPDB:GN00023; CESP:F15H10.2
A;Experimental source: clone F15H10
C;Genetics:
A;Gene: col-13
A;Map position: 5
A;Introns: 71/3
F;1-36/Domain: signal sequence #status predicted <SIG>
F;37-316/Product: collagen col-13 #status predicted <MAT>

Query Match      87.8%; Score 43; DB 2; Length 316;
Best Local Similarity 77.8%; Pred. No. 15;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGSPGSPGQ 9
Db 190 SGAPGPGQ 198

RESULT 10
T22827
hypothetical protein F57B1.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T22827
R;Sims, M.
submitted to the EMBL Data Library, August 1996
A;Reference number: Z19622
A;Accession: T22827
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-356 <WIL>
A;Cross-references: UNIPROT:Q20921; EMBL:Z78064; PIDN:CAB01508.1; GSPDB:GN00023; CESP:F5
A;Experimental source: clone F57B1
C;Genetics:
A;Gene: CESP:F57B1.4
A;Map position: 5
A;Introns: 65/1; 111/3

Query Match      87.8%; Score 43; DB 2; Length 356;
Best Local Similarity 77.8%; Pred. No. 17;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGSPGSPGQ 9
Db 230 SGAPGPGQ 238

RESULT 11
T70893
hypothetical glycine-rich protein Rv1068c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Jul-2004
C;Accession: B70893
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;
Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.;
Nature 393, 537-544, 1998
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: B70893
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-463 <COL>
A;Cross-references: UNIPROT:O53416; GB:AL021897; GB:AL123456; NID:g3256022; PIDN:CAA1718
A;Experimental source: strain H37RV
C;Genetics:
```

A:Gene: Rv1068c

C:Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Query Match 87.8%; Score 43; DB 2; Length 463;

Best Local Similarity 77.8%; Pred. No. 22; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGSPGSPGQ 9

Db 455 SGNPGTGGQ 463

RESULT 12

T15245

hypothetical protein T05E7.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004

C:Accession: T15245

R:Rohlfing, T.; Wohldmann, P.; Biewald, T.

submitted to the EMBL Data Library, May 1997

A:Description: The sequence of C. elegans cosmid T05E7.

A:Reference number: Z18315

A:Accession: T15245

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-208 <ROH>

A:Cross-references: UNIPROT:O01860; EMBL:AF003150; NID:g2088791; PID:g2088793; PIDN:AA85

A:Experimental source: strain Bristol N2; clone T05E7

C:Genetics:

A:Gene: CESP:T05E7.2

A:Map position: 1

A:Introns: 26/2; 62/3

Query Match

Best Local Similarity 85.7%; Score 42; DB 2; Length 208;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SGSPGSPGQ 9

Db 152 SGAPGQPGQ 160

RESULT 13

S34665

collagen, cuticular - root-knot nematode (Meloiodogyne incognita)

C:Species: Meloiodogyne incognita

C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Jul-2004

C:Accession: S34665

R:van der Eycken, W.V.; de Almeida Engler, J.; van Montagu, M.; Gheysen, G.

submitted to the EMBL Data Library, July 1993

A:Description: Identification and analysis of a cuticular collagen gene from the plant-

A:Reference number: S34665

A:Accession: S34665

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-286 <VAN>

A:Cross-references: UNIPROT:Q25467; EMBL:Z24734; NID:g395144; PIDN:CAA80860.1; PID:g3951

Query Match

Best Local Similarity 85.7%; Score 42; DB 2; Length 286;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SGSPGSPGQ 9

Db 235 SGKPGAPGQ 243

RESULT 14

T22639

hypothetical protein F54D1.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C:Accession: T22639

R:Lennard, N.

submitted to the EMBL Data Library, July 1996

A:Reference number: Z19592

A:Accession: T22639

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-294 <WIL>

A:Cross-references: UNIPROT:Q20761; EMBL:Z77132; PIDN:CAB00858.1; GSPDB:GN000022; CESP:F54

A:Experimental source: clone F54D1

C:Genetics:

A:Gene: CESP:F54D1.2

A:Map position: 4

A:Introns: 37/3

Query Match

Best Local Similarity 85.7%; Score 42; DB 2; Length 294;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SGSPGSPGQ 9

Db 167 SGQPGTGGQ 175

RESULT 15

T24482

hypothetical protein T05A1.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C:Accession: T24482

R:Rilloyd, C.

submitted to the EMBL Data Library, December 1995

A:Reference number: Z19897

A:Accession: T24482

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-300 <WIL>

A:Cross-references: UNIPROT:Q22183; EMBL:Z68219; PIDN:CAA92476.1; GSPDB:GN000022; CESP:T0

A:Experimental source: clone T05A1

C:Genetics:

A:Gene: CESP:T05A1.2

A:Map position: 4

A:Introns: 7/3

Query Match

Best Local Similarity 85.7%; Score 42; DB 2; Length 300;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 SGSPGSPGQ 9

Db 241 GAPGSPGQ 248

Search completed: July 18, 2005, 13:41:50

Job time : 14.86 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2005, 12:59:21 ; Search time 68.22 Seconds
(without alignments)
67.557 Million cell updates/sec

Title: SEQ3
Perfect score: 49
Sequence: 1 sgspgspgq 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Uniprot 03.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	46	93.9	306	2	Q20135	Q20135 caenorhabdi
2	45	91.8	1464	2	Q6P912	Q6P912 xenopus lae
3	44	89.8	639	1	P2B1 CRYNV	O42773 cryptococcu
4	44	89.8	1096	1	PUL4 KLEAE	P07811 klebsiella
5	43	87.8	266	1	YXWK CAEEL	Q21184 caenorhabdi
6	43	87.8	299	2	Q20805	Q20805 caenorhabdi
7	43	87.8	299	2	Q22393	Q22393 caenorhabdi
8	43	87.8	302	2	O44904	O44904 caenorhabdi
9	43	87.8	304	2	Q9XUE9	Q9XUE9 caenorhabdi
10	43	87.8	316	1	CC12 CAEEL	P20630 caenorhabdi
11	43	87.8	316	1	CC13 CAEEL	P20631 caenorhabdi
12	43	87.8	316	2	P90728	P90728 caenorhabdi
13	43	87.8	317	2	Q20921	Q20921 caenorhabdi
14	43	87.8	375	2	Q9NAE0	Q9NAE0 caenorhabdi
15	43	87.8	463	1	PG20 MYCTU	O53416 mycobacteri
16	43	87.8	507	2	Q829B3	Q829B3 streptomyce
17	43	87.8	763	2	Q7UOR0	Q7UOR0 mycobacteri
18	43	87.8	854	2	Q8IVT9	Q8IVT9 homo sapien
19	43	87.8	954	2	Q8WVX8	Q8WVX8 homo sapien
20	43	87.8	957	2	Q96P44	Q96P44 homo sapien
21	43	87.8	957	2	Q9H0V3	Q9H0V3 homo sapien
22	43	87.8	1378	2	O97405	O97405 haliotis di
23	43	87.8	1626	2	O8NFW1	O8NFW1 homo sapien
24	42	85.7	208	2	O01860	O01860 caenorhabdi
25	42	85.7	286	2	Q7JMU0	Q7JMU0 meloidogyne
26	42	85.7	294	2	Q20761	Q20761 caenorhabdi
27	42	85.7	300	2	Q22183	Q22183 caenorhabdi
28	42	85.7	305	2	Q25467	Q25467 meloidogyne
29	42	85.7	308	2	O94620	O94620 meloidogyne
30	42	85.7	309	2	Q25466	Q25466 meloidogyne
31	42	85.7	310	2	Q20764	Q20764 caenorhabdi

32	42	85.7	317	2	O9VIX8	O9VIX8 drosophila
33	42	85.7	323	1	CC39 CAEEL	Q09455 caenorhabdi
34	42	85.7	364	2	Q20744	Q20744 caenorhabdi
35	42	85.7	373	2	Q821I1	Q821I1 streptomyce
36	42	85.7	589	2	Q99LL6	Q99LL6 mus musculu
37	42	85.7	599	2	Q00559	Q00559 homo sapien
38	42	85.7	604	2	Q9L252	Q9L252 streptomyce
39	42	85.7	718	2	Q05850	Q05850 homo sapien
40	42	85.7	957	2	Q641F3	Q641F3 xenopus lae
41	42	85.7	1143	1	CALI HUMAN	Q14993 homo sapien
42	42	85.7	1225	2	Q6PCL3	Q6PCL3 mus musculu
43	42	85.7	1453	1	CALI MOUSE	P11087 mus musculu
44	42	85.7	1453	2	Q63079	Q63079 rattus norv
45	42	85.7	1453	2	Q810J9	Q810J9 mus musculu

ALIGNMENTS

RESULT 1
Q20135 PRELIMINARY; PRT; 306 AA.
ID Q20135
AC Q20135;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein F38A3.1.
GN ORFNames=F38A3.1;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditicoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
investigating biology."
RL Science 282:2012-2018 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Swinburne J.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z49938; CAA90187.1; -;
DR PIR; T21938; T21938.
DR WormBase; WEGene0000657; F38A3.1.
DR WormPep; F38A3.1; CE02213.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0042302; F:structural constituent of cuticle; IEA.
DR GO; GO:006817; P:phosphate transport; IEA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR002486; Col_cuticle_N.
DR Pfam; PF01391; Collagen; 3.
DR Pfam; PF01484; Col_cuticle_N; 1.
KW Collagen; Hypothetical protein.
SQ SEQUENCE 306 AA; 30893 MW; 99282F89782796A1 CRC64;

Query Match 93.9%; Score 46; DB 2; Length 306;
Best Local Similarity 88.9%; Pred. No. 47;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 SGSPGSPGQ 9
Db 136 SGNFSGPQ 144
RESULT 2
Q6P912 PRELIMINARY; PRT; 1464 AA.
ID Q6P912
AC Q6P912;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)

05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
MG69046 protein.
Name=MG69046;
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
OC Xenopodinae; Xenopus.
NCBI_TaxID=8355;
[1]
SEQUENCE FROM N.A.
TISSUE=Spleen;
MEDLINE=2238857; PubMed=12477932; DOI=10.1073/pnas.242603899;
Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaney S.J.,
Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Fahey J., Helton E., Kettman M., Madan A.C., Rodrigues S., Sanchez A.,
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
Krzyszynski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
Jones S.J., Marra M.A.;
"Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences";
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[2]
SEQUENCE FROM N.A.
TISSUE=Spleen;
MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
Richardson P.;
"Genetic and genomic tools for Xenopus research: The NIH Xenopus
initiative";
Dev. Dyn. 225:384-391(2002).
[3]
SEQUENCE FROM N.A.
TISSUE=Spleen;
Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.
EMBL; BC060753; AH60753.1; -;
GO; GO:0005581; C:collagen; IEA.
GO; GO:0005731; C:cytoplasm; IEA.
GO; GO:0005203; F:extracellular matrix structural constituent; IEA.
GO; GO:006817; P:phosphate transport; IEA.
InterPro; IPR008161; Clg helix.
InterPro; IPR008160; Collagen.
InterPro; IPR000895; Fib collagen_C.
InterPro; IPR009041; PMP SGGI.
InterPro; IPR001007; VWF_C.
Pfam; PF01410; COLFI; 1.
Pfam; PF01391; Collagen; 18.
Pfam; PF00093; VWC; 1.
ProDom; PD000007; Clg helix; 5.
ProDom; PD002078; Fib collagen_C; 1.
SMART; SM00038; COLFI; 1.
SMART; SM00214; VWC; 1.
PROSITE; PS01208; VWF_C_1; UNKNOWN_1.
PROSITE; PS50184; VWF_C_2; 1.
Collagen.
KW
SEQUENCE 1464 AA; 138840 MW; 0AAA32BE91FBCAF9 CRC64;
Query Match 91.8%; Score 45; DB 2; Length 1464;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
MG69046 protein.
Name=MG69046;
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
OC Xenopodinae; Xenopus.
NCBI_TaxID=8355;
[1]
SEQUENCE FROM N.A.
TISSUE=Spleen;
MEDLINE=2238857; PubMed=12477932; DOI=10.1073/pnas.242603899;
Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaney S.J.,
Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Fahey J., Helton E., Kettman M., Madan A.C., Rodrigues S., Sanchez A.,
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
Krzyszynski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
Jones S.J., Marra M.A.;
"Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences";
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[2]
SEQUENCE FROM N.A.
TISSUE=Spleen;
MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
Richardson P.;
"Genetic and genomic tools for Xenopus research: The NIH Xenopus
initiative";
Dev. Dyn. 225:384-391(2002).
[3]
SEQUENCE FROM N.A.
TISSUE=Spleen;
Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.
EMBL; BC060753; AH60753.1; -;
GO; GO:0005581; C:collagen; IEA.
GO; GO:0005731; C:cytoplasm; IEA.
GO; GO:0005203; F:extracellular matrix structural constituent; IEA.
GO; GO:006817; P:phosphate transport; IEA.
InterPro; IPR008161; Clg helix.
InterPro; IPR008160; Collagen.
InterPro; IPR000895; Fib collagen_C.
InterPro; IPR009041; PMP SGGI.
InterPro; IPR001007; VWF_C.
Pfam; PF01410; COLFI; 1.
Pfam; PF01391; Collagen; 18.
Pfam; PF00093; VWC; 1.
ProDom; PD000007; Clg helix; 5.
ProDom; PD002078; Fib collagen_C; 1.
SMART; SM00038; COLFI; 1.
SMART; SM00214; VWC; 1.
PROSITE; PS01208; VWF_C_1; UNKNOWN_1.
PROSITE; PS50184; VWF_C_2; 1.
Collagen.
KW
SEQUENCE 1464 AA; 138840 MW; 0AAA32BE91FBCAF9 CRC64;
Query Match 91.8%; Score 45; DB 2; Length 1464;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
MG69046 protein.
Name=MG69046;
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
OC Xenopodinae; Xenopus.
NCBI_TaxID=8355;
[1]
SEQUENCE FROM N.A.
TISSUE=Spleen;
MEDLINE=2238857; PubMed=12477932; DOI=10.1073/pnas.242603899;
Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaney S.J.,
Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Fahey J., Helton E., Kettman M., Madan A.C., Rodrigues S., Sanchez A.,
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
Krzyszynski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
Jones S.J., Marra M.A.;
"Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences";
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[2]
SEQUENCE FROM N.A.
TISSUE=Spleen;
MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
Richardson P.;
"Genetic and genomic tools for Xenopus research: The NIH Xenopus
initiative";
Dev. Dyn. 225:384-391(2002).
[3]
SEQUENCE FROM N.A.
TISSUE=Spleen;
Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.
EMBL; BC060753; AH60753.1;

```

Db      554 SGSPGSPG 561

RESULT 4
PULA_KLEAE
ID      PULA_KLEAE      STANDARD;      PRT;      1096 AA.
AC      P07811;
DT      01-AUG-1988 (Rel. 08, Created)
DT      01-AUG-1988 (Rel. 08, Last sequence update)
DT      25-OCT-2004 (Rel. 45, Last annotation update)
DE      Pullulanase precursor (EC 3.2.1.41) (Alpha-dextrin endo-1,6-alpha-
DE      glucosidase) (Pullulan 6-glucanohydrolase).
GN      Namespula;
OS      Klebsiella aerogenes.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC      Enterobacteriaceae; Klebsiella.
OX      NCBI_TaxID=28451;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=W70;
RX      MEDLINE=87194626; PubMed=3155373;
RA      Katsuragi N., Takizawa N., Murooka Y.;
RT      "Entire nucleotide sequence of the pullulanase gene of Klebsiella
RT      aerogenes W70.";
RL      J. Bacteriol. 169:2301-2306(1987).
CC      -1- CATALYTIC ACTIVITY: Hydrolysis of (1->6)-alpha-D-glucosidic
CC      linkages in pullulan and in amylopectin and glycogen, and the
CC      alpha- and beta-limit dextrins of amylopectin and glycogen.
CC      -1- SUBUNIT: Homotrimer.
CC      -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC      (Probable).
CC      -1- SIMILARITY: Belongs to the glycosyl hydrolase 13 family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; M16187; AAA25124.1; ALT_SEQ.
DR      PIR; A26879; A26879.
DR      HSP; P07762; IM7X.
DR      InterPro; IPR006047; Alpha_amyl_cat.
DR      InterPro; IPR004193; Glyco_hydro_13N.
DR      InterPro; IPR005323; PUD.
DR      Pfam; PF00128; Alpha-amylase; 1.
DR      Pfam; PF02922; Isoamylase_N; 1.
DR      Pfam; PF03714; PUD; 1.
DR      PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW      Glycosidase; Hydrolase; Lipoprotein; Membrane; Palmitate; Signal.
FT      SIGNAL      1      19
FT      CHAIN      20      1096
FT      LIPID      20      20
FT      LIPID      20      20
FT      ACT_SITE      694      694
FT      ACT_SITE      723      723
FT      ACT_SITE      851      851
FT      ACT_SITE      851      851
SQ      SEQUENCE      1096 AA; 119335 MW; FE7D9167CDACFD79 CRC64;

Query Match      89.8%; Score 44; DB 1; Length 1096;
Best Local Similarity 100.0%; Pred. NO. 3.4e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 SGSPGSPG 8
Db      29 SGSPGSPG 36

RESULT 5
YXWK_CABEL
ID      YXWK_CABEL      STANDARD;      PRT;      266 AA.

```

```

AC      Q21184; Q20807;
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      25-OCT-2004 (Rel. 45, Last annotation update)
DE      Putative cuticle collagen F55C10.3.
GN      ORFNames=F55C10.3;
OS      Caenorhabditis elegans.
OC      Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC      Rhabditidae; Peloderinae; Caenorhabditis.
OX      NCBI_TaxID=6239;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Bristol N2;
RX      MEDLINE=99089613; PubMed=9851916;
RG      The C. elegans sequencing consortium;
RT      "Genome sequence of the nematode C. elegans: a platform for
RT      investigating biology.";
RL      Science 282:2012-2018(1998).
RN      [2]
RP      REVISIONS.
RA      Jones S.J.M.;
RL      Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC      -1- FUNCTION: Nematode cuticles are composed largely of collagen-like
CC      proteins. The cuticle functions both as an exoskeleton and as a
CC      barrier to protect the worm from its environment (By similarity).
CC      -1- SUBUNIT: Collagen polypeptide chains are complexed within the
CC      cuticle by disulfide bonds and other types of covalent cross-links
CC      (By similarity).
CC      -1- SIMILARITY: Belongs to the cuticular collagen family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; Z74036; CAA398487.1; -.
DR      PIR; T22706; T22706.
DR      WormBase; WBGene0000728; F55C10.3.
DR      WormPep; F55C10.3; CE11182.
DR      InterPro; IPR008160; Collagen.
DR      Pfam; PF01391; Collagen; 2.
KW      Collagen; Cuticle; Hypothetical protein; Multigene family; Repeat;
KW      Structural protein.
FT      DOMAIN      70      99
FT      DOMAIN      118      144
FT      DOMAIN      148      169
FT      DOMAIN      183      245
SQ      SEQUENCE      266 AA; 25616 MW; 984556680F1AAF22 CRC64;

Query Match      87.8%; Score 43; DB 1; Length 266;
Best Local Similarity 77.8%; Pred. NO. 1.1e+02;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 SGSPGSPG 9
Db      135 NGNPGSPG 143

RESULT 6
Q20805
ID      Q20805      PRELIMINARY;      PRT;      299 AA.
AC      Q20805;
DT      01-NOV-1996 (TrEMBLrel. 01, Created)
DT      01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT      01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE      Hypothetical protein F55C10.2.
GN      ORFNames=F55C10.2;
OS      Caenorhabditis elegans.
OC      Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC      Rhabditidae; Peloderinae; Caenorhabditis.

```

```

OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Dobson R.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z74036; CAA98486.1; -.
DR FIR; T22705; T22705.
DR WormBase; WBGene00000727; F55C10.2.
DR WormPep; F55C10.2; CE05952.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0042302; F:structural constituent of cuticle; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR002486; Col_cuticle_N.
DR Pfam; PF01391; Collagen; 2.
DR Pfam; PF01484; Col_cuticle_N; 1.
KW Collagen; Hypothetical protein.
SQ SEQUENCE 299 AA; 29280 MW; 59F864183BE937B2 CRC64;

Query Match 87.8%; Score 43; DB 2; Length 299;
Best Local Similarity 77.8%; Pred. NO. 1.3e+02;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SQSPGSPGQ 9
Db 168 NGNPGSPGQ 176

RESULT 7
Q22393 PRELIMINARY; PRT; 299 AA.
AC Q22393;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein T11F9.9.
GN ORFNames=T11F9.9;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Lennard N.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z74042; CAA98525.1; -.
DR FIR; T24833; T24833.
DR WormBase; WBGene00000730; T11F9.9.
DR WormPep; T11F9.9; CE06421.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0042302; F:structural constituent of cuticle; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR002486; Col_cuticle_N.
DR Pfam; PF01391; Collagen; 2.

QY 1 SQSPGSPGQ 9
Db 168 NGNPGSPGQ 176

RESULT 7
Q22393 PRELIMINARY; PRT; 299 AA.
AC Q22393;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein T11F9.9.
GN ORFNames=T11F9.9;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Lennard N.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z74042; CAA98525.1; -.
DR FIR; T24833; T24833.
DR WormBase; WBGene00000730; T11F9.9.
DR WormPep; T11F9.9; CE06421.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0042302; F:structural constituent of cuticle; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR002486; Col_cuticle_N.
DR Pfam; PF01391; Collagen; 2.

QY 1 SQSPGSPGQ 9
Db 168 NGNPGSPGQ 176

RESULT 7
Q22393 PRELIMINARY; PRT; 299 AA.
AC Q22393;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein T11F9.9.
GN ORFNames=T11F9.9;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: a platform for
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA Davidson S., Langston Y., O'Neal D.;
RT "The sequence of C. elegans cosmid W05G11.1";
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX WormBase Consortium;
RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF040660; AAC71147.1; -.
DR FIR; T32872; T32872.
DR WormBase; WBGene00000663; W05G11.3.
DR WormPep; W05G11.3; CE14628.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0042302; F:structural constituent of cuticle; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR002486; Col_cuticle_N.
DR Pfam; PF01391; Collagen; 3.
DR Pfam; PF01484; Col_cuticle_N; 1.
KW Collagen.
SQ SEQUENCE 302 AA; 29112 MW; E3BBFA74248A597F CRC64;

Query Match 87.8%; Score 43; DB 2; Length 302;
Best Local Similarity 77.8%; Pred. NO. 1.3e+02;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SQSPGSPGQ 9
Db 162 AGSPGAPGQ 170

```

```

DR Pfam; PF01484; Col_cuticle_N; 1.
KW Collagen; Hypothetical protein.
SQ SEQUENCE 299 AA; 29269 MW; 86AC5101AD0A150B CRC64;

Query Match 87.8%; Score 43; DB 2; Length 299;
Best Local Similarity 77.8%; Pred. NO. 1.3e+02;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SQSPGSPGQ 9
Db 168 NGNPGSPGQ 176

RESULT 8
O44904 PRELIMINARY; PRT; 302 AA.
ID O44904;
AC O44904;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Collagen protein 88.
GN Name=col-88; ORFNames=W05G11.3;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RG WormBase Consortium;
RT "Genome sequence of the nematode C. elegans: a platform for
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Davidson S., Langston Y., O'Neal D.;
RT "The sequence of C. elegans cosmid W05G11.1";
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX WormBase Consortium;
RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF040660; AAC71147.1; -.
DR FIR; T32872; T32872.
DR WormBase; WBGene00000663; W05G11.3.
DR WormPep; W05G11.3; CE14628.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0042302; F:structural constituent of cuticle; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR002486; Col_cuticle_N.
DR Pfam; PF01391; Collagen; 3.
DR Pfam; PF01484; Col_cuticle_N; 1.
KW Collagen.
SQ SEQUENCE 302 AA; 29112 MW; E3BBFA74248A597F CRC64;

Query Match 87.8%; Score 43; DB 2; Length 302;
Best Local Similarity 77.8%; Pred. NO. 1.3e+02;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SQSPGSPGQ 9
Db 162 AGSPGAPGQ 170

```

```

RESULT 9
Q9XUE9 QXUE9 PRELIMINARY; PRT; 304 AA.
AC Q9XUE9;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein F52B11.4.
GN ORFNames=F52B11.4;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Matthews L.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z82268; CAB05195.1; -.
DR PIR; T22482; T22482.
DR WormBase; WBGene0000707; F52B11.4.
DR WormPep; F52B11.4; CE18724.
DR GO; GO:0005737; Cytoplasm; IEA.
DR GO; GO:0042302; F:structural constituent of cuticle; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR002486; Col cuticle_N.
DR Pfam; PF01391; Collagen; 2.
DR Pfam; PF01484; Col cuticle_N; 1.
KW Collagen; Hypothetical protein.
SQ SEQUENCE 304 AA; 29144 MW; 68C73E2551E72D8A CRC64;

Query Match 87.8%; Score 43; DB 2; Length 304;
Best Local Similarity 77.8%; Pred. No. 1.3e+02;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGSPGSPGQ 9
Db 194 SGAPGPGQ 202
||:|:|:|

RESULT 10
CC12_CAEEL
ID CC12_CAEEL STANDARD; PRT; 316 AA.
AC P20630;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Cuticle collagen 12 precursor.
GN Names-col-12; ORFNames=F15H10.1;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=90172409; PubMed=1689778;
RA Park Y.-S., Kramer J.M.;
RT "Tandemly duplicated Caenorhabditis elegans collagen genes differ in
RT their modes of splicing.";
RL J. Mol. Biol. 211:395-406(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;

Query Match 87.8%; Score 43; DB 1; Length 316;
Best Local Similarity 77.8%; Pred. No. 1.3e+02;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGSPGSPGQ 9
Db 190 SGAPGPGQ 198
||:|:|:|

RESULT 11
CC13_CAEEL
ID CC13_CAEEL STANDARD; PRT; 316 AA.
AC P20631;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Cuticle collagen 13 precursor.
GN Names-col-13; ORFNames=F15H10.2;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=90172409; PubMed=1689778;
RA Park Y.-S., Kramer J.M.;
RT "Tandemly duplicated Caenorhabditis elegans collagen genes differ in
RT their modes of splicing.";
RL J. Mol. Biol. 211:395-406(1990).
RN [2]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=99069613; PubMed=9851916;
RT The C. elegans sequencing consortium;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
CC -!- FUNCTION: Nematode cuticles are composed largely of collagen-like
CC proteins. The cuticle functions both as an exoskeleton and as a
CC barrier to protect the worm from its environment.
CC -!- SUBUNIT: Collagen polypeptide chains are complexed within the
CC cuticle by disulfide bonds and other types of covalent cross-
CC links.
CC -!- SIMILARITY: Belongs to the cuticular collagen family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X51622; CAA35954.1; -.
CC EMBL; Z73972; CAA98257.1; -.
CC PIR; S08169; S08169.
CC WormBase; WBGene0000601; F15H10.1.
CC WormPep; F15H10.1; CE05638.
CC InterPro; IPR002486; Col cuticle_N.
CC InterPro; IPR008160; Collagen.
CC Pfam; PF01484; Col cuticle_N; 1.
CC Pfam; PF01391; Collagen; 3.
KW Collagen; Cuticle; Multigene family; Repeat; Signal;
KW Structural protein.
FT SIGNAL 1 36 Potential.
FT CHAIN 37 316 Cuticle collagen 12.
FT DOMAIN 128 157 Triple-helical region.
FT DOMAIN 176 202 Triple-helical region.
FT DOMAIN 206 235 Triple-helical region.
FT DOMAIN 240 266 Triple-helical region.
FT DOMAIN 269 304 Triple-helical region.
SQ SEQUENCE 316 AA; 30098 MW; 6CA81FF94706D42E CRC64;

Query Match 87.8%; Score 43; DB 1; Length 316;
Best Local Similarity 77.8%; Pred. No. 1.3e+02;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGSPGSPGQ 9
Db 190 SGAPGPGQ 198
||:|:|:|

```

```

RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RG The C. elegans sequencing consortium;
RT "Genome sequence of the nematode C. elegans: a platform for
  investigating biology.";
RL Science 282:2012-2018(1998).
CC -!- FUNCTION: Nematode cuticles are composed largely of collagen-like
CC proteins. The cuticle functions both as an exoskeleton and as a
CC barrier to protect the worm from its environment.
CC -!- SUBUNIT: Collagen polypeptide chains are complexed within the
CC cuticle by disulfide bonds and other types of covalent cross-
CC links.
CC -!- SIMILARITY: Belongs to the cuticular collagen family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X51623; CAA35955.1; -.
DR EMBL; Z73972; CAA398258.1; -.
DR FIR; S08170; S08170.
DR WormBase; WBGene0000602; F15H10.2.
DR WormPep; F15H10.2; CE05639.
DR InterPro; IPR002486; Col_cuticle_N.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF01484; Col_cuticle_N; 1.
DR Pfam; PF01391; Collagen; 3.
DR Collagen; Cuticle; Multigene family; Repeat; Signal;
KW Structural protein..
FT SIGNAL 1 36 Potential.
FT CHAIN 37 316 Cuticle collagen 13.
FT DOMAIN 128 157 Triple-helical region.
FT DOMAIN 176 202 Triple-helical region.
FT DOMAIN 206 235 Triple-helical region.
FT DOMAIN 240 266 Triple-helical region.
FT DOMAIN 269 304 Triple-helical region.
SQ SEQUENCE 316 AA; 30100 MW; 00C6D08FBC4701AF CRC64;

  Query Match      87.8%; Score 43; DB 1; Length 316;
  Best Local Similarity 77.8%; Pred. No. 1.3e+02;
  Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGSPGSPGQ 9
Db 190 SGAPGAPGQ 198

RESULT 12
ID P90728 PRELIMINARY; PRT; 316 AA.
AC P90728;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Cuticle collagen.
GN Names=Ccol-12;
OS Caenorhabditis briggsae.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6238;
RN [1]
RP SEQUENCE FROM N.A.
RA Gilleard J.S., Henderson D.K., Ulla N.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U84501; ABA41793.1; -.
DR GO; GO:0005737; Cytoplasm; IEA.
DR GO; GO:0042302; F:structural constituent of cuticle; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008160; Collagen.

  Query Match      87.8%; Score 43; DB 1; Length 316;
  Best Local Similarity 77.8%; Pred. No. 1.3e+02;
  Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGSPGSPGQ 9
Db 190 SGAPGAPGQ 198

RESULT 13
ID Q20921 PRELIMINARY; PRT; 317 AA.
AC Q20921;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein F57B1.4.
GN ORFNames=F57B1.4;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA none;
RX STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
  investigating biology.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Sims M.A.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z78064; CAB01508.2; -.
DR FIR; T22827; T22827.
DR WormBase; WBGene0000733; F57B1.4.
DR WormPep; F57B1.4; CE31962.
DR GO; GO:0005737; Cytoplasm; IEA.
DR GO; GO:0042302; F:structural constituent of cuticle; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR002486; Col_cuticle_N.
DR Pfam; PF01391; Collagen_3.
DR Pfam; PF01484; Col_cuticle_N; 1.
DR Collagen; Hypothetical protein.
KW SEQUENCE 317 AA; 30281 MW; 4E6BA587AF105A8D CRC64;

  Query Match      87.8%; Score 43; DB 2; Length 317;
  Best Local Similarity 77.8%; Pred. No. 1.3e+02;
  Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGSPGSPGQ 9
Db 191 SGAPGAPGQ 199

RESULT 14
ID Q9NAE0 PRELIMINARY; PRT; 375 AA.
AC Q9NAE0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein Y51H4A.9.
GN ORFNames=Y51H4A.9;

```

```

DR InterPro; IPR002486; Col_cuticle_N.
DR Pfam; PF01391; Collagen; 3.
DR Pfam; PF01484; Col_cuticle_N; 1.
KW Collagen.
SQ SEQUENCE 316 AA; 30045 MW; D36639D5E9EF2243 CRC64;

  Query Match      87.8%; Score 43; DB 2; Length 316;
  Best Local Similarity 77.8%; Pred. No. 1.3e+02;
  Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGSPGSPGQ 9
Db 190 SGAPGAPGQ 198

```

```

OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=Bristol N2;
RX MEDLINE=99089613; PubMed=9851916;
RT none;
RA "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
RN [2]
RC SEQUENCE FROM N.A.
RP STRAIN=Bristol N2;
RA Sulston J.E.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL132952; CAB61143.1; -.
DR WormBase: WBGene0000710; Y51H4A.9.
DR WormRep: Y51H4A.9; CE22336.
DR GO: GO:0005737; C:cytoplasm; IEA.
DR GO: GO:0042302; F:structural constituent of cuticle; IEA.
DR GO: GO:0006817; P:phosphate transport; IEA.
DR InterPro: IPR008160; Collagen.
DR InterPro: IPR002486; Col_cuticle_N.
DR Pfam: PF01391; Collagen; 2.
DR Pfam: PF01484; Col_cuticle_N; 1.
KW Collagen, Hypothetical protein.
SQ SEQUENCE 375 AA; 40765 MW; 49C1AED2E10CD356 CRC64;

Query Match 87.8%; Score 43; DB 2; Length 375;
Best Local Similarity 77.8%; Pred. No. 1.6e+02;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGSPGSPGQ 9
DB 225 SGAPGAPGQ 233

RESULT 15
PG20 MYCTU STANDARD; PRT; 463 AA.
AC OS3416;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 25-JAN-2005 (Rel. 46, Last annotation update)
DE Hypothetical PE-PGRS family protein PE_PGRS20.
GN Name=PE_PGRS20; OrderedLocustNames=Rv1068c, MT1097;
GN ORFNames=MTV017.21c;
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230; DOI=10.1038/31159;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C.M.,
RA Harris D.E., Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III,
RA Tekala F., Badcock K., Basham D., Brown D., Chillingworth T.,
RA Connor R., Davies R.M., Devlin K., Feltwell T., Gentles S., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S.,
RA Murphy L.D., Oliver S., Osborne J., Quail M.A., Rajandream M.A.,
RA Rogers J., Rutter S., Seeger K., Skelton S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RC SEQUENCE FROM N.A.
RP STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RX DOI=10.1128/JB.184.19.5479-5490.2002;

```

```

RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J.D., DeBoy R.T., Dodson R.J., Winn M.L., Haft D.H.,
RA Hickey E.K., Kolonay J.F., Nelson W.C., Unayam L.A., Ermolaeva M.D.,
RA Salzberg S.L., Delcher A., Utterback T.R., Weidman J.F., Khouri H.M.,
RA Gill J., Mikula A., Bishai W., Jacobs W.R. Jr., Venter J.C.,
RA Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol. 184:5479-5490(2002).
CC -1- SIMILARITY: Belongs to the mycobacterial PE family. PGRS
CC subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL: BX842575; CAE55352.1; -.
CC DR EMBL: AB000516; AAK45353.1; ALT_INIT.
CC DR PIR: B70893; B70893.
CC DR TIGR: MT1097; -.
CC DR TubercuList; Rv1068c; -.
CC DR InterPro; IPR009050; Globin_like.
CC DR InterPro; IPR000084; PE_region_N.
CC DR Pfam: PF00934; PE; 1.
CC DR ProDom: PD001223; PE_region_N; 1.
CC DR Complete proteome; Hypothetical protein.
KW Complete proteome; Hypothetical protein.
FT CONFLICT 218 218 T -> S (in Ref. 2).
FT CONFLICT 235 235 G -> GGGGAGIGGADTKGGDAGAGAGAGAGGWHGHGV
FT GDDGGTGGGGGQGGVQGGPDGTGAAGGAGG (in Ref.
FT 2).
SQ SEQUENCE 463 AA; 39305 MW; CF5696A7E9593952 CRC64;

Query Match 87.8%; Score 43; DB 1; Length 463;
Best Local Similarity 77.8%; Pred. No. 2e+02;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGSPGSPGQ 9
DB 455 SGNPGTPGQ 463

Search completed: July 18, 2005, 13:33:21
Job time : 70.22 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2005, 12:59:21 ; Search time 75.78 Seconds
(without alignments)
45.934 Million cell updates/sec

Title: SEQ3
Perfect score: 49
Sequence: 1 sgspgpggq 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	49	100.0	9	3 AAY71088	Synthetic
2	44	89.8	1096	1 AAP82507	Abp82507 Pullulana
3	44	89.8	1111	7 ABO65813	Abp65813 Klebsiell
4	43	87.8	107	5 ABP42940	Abp42940 Human ova
5	43	87.8	429	4 AAU16066	Aau16066 Human nov
6	43	87.8	429	6 ABUS5135	Abus5135 Human nov
7	43	87.8	957	4 ABUS2683	Abus2683 Human cel
8	43	87.8	1606	6 ABR40002	Abp40002 Human col
9	43	87.8	1626	6 ABR40001	Abp40001 Human col
10	42	85.7	104	3 AAB06022	Abp06022 Polar gel
11	42	85.7	161	5 ABG31593	Abg31593 Synthetic
12	42	85.7	161	5 ABB98499	Abb98499 Mature H-
13	42	85.7	317	4 ABB67768	Abb67768 Drosophil
14	42	85.7	401	3 AAB06023	Abp06023 Polar gel
15	42	85.7	599	4 AAY72375	Aay72375 Amphiphil
16	42	85.7	599	4 AAY72374	Aay72374 Amphiphil
17	42	85.7	1142	8 ADQ59442	Adq59442 Human can
18	42	85.7	1142	8 ABM82527	Abm82527 Human dia
19	42	85.7	1142	8 ABM82528	Abm82528 Human dia
20	42	85.7	1453	5 ABG393948	Abg393948 Mouse pol
21	42	85.7	1453	7 ADD45053	Add45053 Rat Prote
22	42	85.7	1453	7 ADD45057	Add45057 Rat Prote
23	42	85.7	1453	7 ADD48341	Add48341 Rat Prote
24	42	85.7	1453	7 ADD45049	Add45049 Rat Prote
25	42	85.7	1453	7 ADD48337	Add48337 Rat Prote

26	42	85.7	1453	7 ADD48345	Add48345 Rat Prote
27	42	85.7	1665	8 ABM84470	Abm84470 Human dia
28	42	85.7	1678	8 ADQ21326	Adq21326 Human sof
29	42	85.7	1694	2 AAW40109	Aaw40109 Human alp
30	42	85.7	1744	8 ADN23179	Adn23179 Bacterial
31	42	85.7	1759	8 ADN24379	Adn24379 Bacterial
32	42	85.7	1779	4 ABB60207	Abb60207 Drosophil
33	41	83.7	21	4 AAB85621	Aab85621 Specifici
34	41	83.7	21	4 AAB85623	Aab85623 Specifici
35	41	83.7	21	4 AAB85625	Aab85625 Specifici
36	41	83.7	21	4 AAB85624	Aab85624 Specifici
37	41	83.7	21	4 AAB85622	Aab85622 Specifici
38	41	83.7	21	4 AAB85626	Aab85626 Specifici
39	41	83.7	59	4 AAB68058	Aab68058 Amino aci
40	41	83.7	59	4 AAE02704	Aae02704 Human alp
41	41	83.7	59	7 ADB84291	Adb84291 Recombina
42	41	83.7	101	4 AAB68059	Aab68059 Amino aci
43	41	83.7	101	4 AAE02705	Aae02705 Human alp
44	41	83.7	101	7 ADB84292	Adb84292 Recombina
45	41	83.7	113	5 ADK34259	Adk34259 Novel hum

ALIGNMENTS

RESULT 1
AAY71088
ID AAY71088 standard; peptide; 9 AA.
XX
AC AAY71088;
XX
DT 21-SEP-2000 (first entry)
DE
DE Synthetic linker peptide #3 encoded by MV05JA oligonucleotide linker.
XX
KW llama; HC-V; heavy chain variable domain; antigen binding protein;
KW linker; conformational flexibility; multivalent binding protein; bi-head;
KW human chorionic gonadotropin; hCG; immunoassay; agglutination assay;
purification.
XX
OS Synthetic.
XX
PH Key
FT Peptide
FT /label= Peptide linker 3
FT /note= "Planked by one residue from N- and C-terminii of
FT HCV fragment"
XX
WO200024884-A2.
XX
04-MAY-2000.
XX
22-OCT-1999; 99WO-EP008323.
XX
27-OCT-1998; 98WO-EP006991.
XX
22-APR-1999; 99EP-00303118.
(UNIL) UNILEVER PLC.
(UNIL) UNILEVER NV.
(HIND-) HINDUSTAN LEVER LTD.
Frenken LGJ, Howell S, Van Der Vaart JM;
WPI; 2000-350728/30.
N-PSDB; AAD00661.
Use of a linker whose amino acid sequence confers restricted
conformational flexibility to generate multivalent and multispecific
antigen binding proteins.
Example 1.1d; Page 20; 50pp; English.
The present sequence is the synthetic linker peptide #3, encoded by the

CC oligonucleotide linker fragment, MV05JA. It consists of the last residue
 CC of the N-terminal HC-V fragment (S) and the first residue of the C-
 CC terminal HC-V fragment (O), intersected by the connecting linker peptide.
 CC It is used for the construction of *Saccharomyces cerevisiae* episomal
 CC expression plasmid, pUR5332, encoding anti-hCG-anti-RR6 bispecific
 CC biheads, containing the linker peptide. The peptide linker confers
 CC restricted conformational flexibility for linking binding units in a
 CC multivalent binding protein. The linker is used to generate multivalent
 CC or multispecific antigen binding proteins for immunoassays, agglutination
 CC assays or for purification
 XX
 XX Sequence 9 AA;

Query Match 100.0%; Score 49; DB 3; Length 9;
 Best Local Similarity 100.0%; Pred. No. 1.8e+06;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGSPGSPGQ 9
 DB 1 SGSPGSPGQ 9

RESULT 2
 ID AAP82507 standard; protein; 1096 AA.

XX AAP82507;

XX 25-MAR-2003 (revised)

DT 01-NOV-1990 (first entry)

XX Pullulanase protein.

XX Pullulanase; starch; alcohol prodn.

XX *Klebsiella* aerogenes.

XX JP63245676-A.

PN 12-OCT-1988.

XX 31-MAR-1987; 87JP-00078355.

XX 31-MAR-1987; 87JP-00078355.

XX (SUNR) SUNTORY LTD.

PA (ELED) DENKI KAGAKU KOGYO KK.

XX WPI; 1988-333488/47.

DR N-PSDB; AAN81341.

PT Gene encoding pullulanase - derived from recombinant plasmid pMPI contg.
 PT gene from *Klebsiella* genus.

XX Disclosure; Page 7; 12pp; Japanese.

XX The pullulanase protein cleaves alpha-1,6-glucoside bonds of starch and
 CC is effective in decomposition of branched starch. It is used in the
 CC prodn. of maltose and glucose from starch, and of alcohol from starch via
 CC glucose. Amino acid residues 1-19 can be deleted. (Updated on 25-MAR-2003
 CC to correct PA field.)
 XX

XX Sequence 1096 AA;

Query Match 89.8%; Score 44; DB 1; Length 1096;
 Best Local Similarity 100.0%; Pred. No. 4.8e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGSPGSPG 8
 DB 29 SGSPGSPG 36

RESULT 3
 ABO65813
 ID ABO65813 standard; protein; 1111 AA.
 XX
 AC ABO65813;
 XX 29-JUL-2004 (first entry)
 DT
 XX
 DE *Klebsiella pneumoniae* polypeptide seqid 12330.

XX Recombinant expression vector; transcription regulatory element;
 KW *Klebsiella pneumoniae* protein; antibacterial; Vaccine.

OS *Klebsiella pneumoniae*.

XX US6610836-B1.

PN 26-AUG-2003.

PD 27-JAN-2000; 2000US-00489039.

PF 29-JAN-1999; 99US-0117747P.

XX (GENO-) GENOME THERAPEUTICS CORP.

PI Breton GL, Osborne M;

XX WPI; 2003-895346/82.

DR N-PSDB; ACH99364.

XX New nucleic acid encoding a *Klebsiella pneumoniae* polypeptide, useful for
 PT preparing a vaccine composition against *Klebsiella pneumoniae*.

XX Disclosure; SEQ ID NO 12330; 932pp; English.

XX The invention describes a new isolated nucleic acid encoding a *Klebsiella*
 CC *pneumoniae* polypeptide. Also described are: a recombinant expression
 CC vector comprising the nucleic acid, operably linked to a transcription
 CC regulatory element; and a cell comprising the recombinant expression
 CC vector. The nucleic acid is useful for preparing a vaccine composition
 CC against *Klebsiella pneumoniae*. This is the amino acid sequence of a
 CC *Klebsiella pneumoniae* polypeptide of the invention
 XX

SQ Sequence 1111 AA;

Query Match 89.8%; Score 44; DB 7; Length 1111;
 Best Local Similarity 100.0%; Pred. No. 4.9e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGSPGSPG 8
 DB 38 SGSPGSPG 45

RESULT 4
 ABP42940
 ID ABP42940 standard; protein; 107 AA.

XX ABP42940;

XX 22-AUG-2002 (first entry)

DE Human ovarian antigen HPODV37, SEQ ID NO:4072.

XX Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
 KW ovarian cancer; breast cancer; tumour; reproductive system disorder;
 KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
 KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;
 KW inflammatory condition; immune disorder; blood disorder;
 KW cardiovascular disorder; respiratory disorder; neurological disorder;
 KW gastrointestinal disorder; urinary system disorder; drug screening;
 KW gene therapy; chromosome mapping; forensic analysis;
 KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;

KW antiinflammatory; gynaecological; reproductive.
 OS Homo sapiens.
 XX
 PN WO200200677-A1.
 XX
 PD 03-JAN-2002.
 XX
 PF 07-JUN-2001; 2001WO-US018569.
 XX
 PR 07-JUN-2000; 2000US-0209467P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Birse CE, Rosen CA;
 XX
 DR WPI; 2002-147878/19.
 DR N-PSDB; ABQ56017.
 XX
 PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,
 PT useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian
 PT cancer), immune disorders, cardiovascular disorders and neurological
 PT diseases.
 XX
 PS Claim 11; SEQ ID NO 4072; 2922pp; English.
 XX
 CC The invention relates to 2175 novel human ovarian antigens (ABP41054-
 CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
 CC encompasses polypeptides 90% identical and polynucleotides 95% identical
 CC to the sequences of the invention. The invention additionally relates to
 CC recombinant vectors and host cells comprising human ovarian antigen
 CC polynucleotides, antibodies against human ovarian antigens, and the use
 CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
 CC treating, prognosing or preventing various ovary and/or breast-related
 CC disorders. Such conditions include ovarian cancer and breast cancer, and
 CC metastatic tumours of ovarian or breast origin, reproductive system
 CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
 CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
 CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
 CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
 CC vaginitis), immune disorders (e.g., congenital and acquired
 CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
 CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
 CC respiratory disorders, neurological disorders, gastrointestinal disorders
 CC and urinary system disorders. Ovarian antigen polypeptides and
 CC polynucleotides may also be used in screening for compounds which
 CC modulate ovarian antigen expression or activity. The polynucleotides may
 CC further be used for gene therapy, chromosome mapping, in the
 CC identification of individuals and in forensic analysis, and the
 CC polypeptides may be used as food additives or to prepare antibodies
 CC useful in disease diagnosis, drug targeting and phenotyping. The present
 CC sequence represents a human ovarian antigen of the invention. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 107 AA;
 Query Match 87.8%; Score 43; DB 5; Length 107;
 Best Local Similarity 88.9%; Pred. No. 73;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 SGSPGSPGQ 9
 Db 88 SGSPGLPGQ 96
 |||||
 RESULT 5
 ID AAU16066
 ID AAU16066 standard; protein; 429 AA.
 XX
 AC AAU16066;
 XX

DT 07-NOV-2001 (first entry)
 XX Human novel secreted protein, Seq ID 1019.
 DE
 XX Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;
 KW cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;
 KW antibacterial; virucide; fungicide; ophthalmological; vulnerary;
 KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;
 KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
 KW cerebral ischaemia; angiogenesis; nervous system disorder;
 KW Alzheimer's disease; infection; ocular disorder; corneal infection;
 KW wound healing; epithelial cell proliferation; skin ageing; food additive;
 KW preservative; antiproliferative.
 XX
 OS Homo sapiens.
 XX
 PN WO200155322-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US001341.
 XX
 PR 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-0180628P.
 PR 24-FEB-2000; 2000US-0184664P.
 PR 02-MAR-2000; 2000US-0186350P.
 PR 16-MAR-2000; 2000US-0189874P.
 PR 18-APR-2000; 2000US-0190076P.
 PR 17-MAR-2000; 2000US-0190076P.
 PR 19-MAY-2000; 2000US-0205515P.
 PR 07-JUN-2000; 2000US-0209467P.
 PR 28-JUN-2000; 2000US-0214886P.
 PR 30-JUN-2000; 2000US-0215135P.
 PR 07-JUL-2000; 2000US-0216647P.
 PR 07-JUL-2000; 2000US-0216880P.
 PR 11-JUL-2000; 2000US-0217487P.
 PR 11-JUL-2000; 2000US-0217496P.
 PR 14-JUL-2000; 2000US-0218290P.
 PR 26-JUL-2000; 2000US-0220963P.
 PR 26-JUL-2000; 2000US-0220964P.
 PR 14-AUG-2000; 2000US-0224518P.
 PR 14-AUG-2000; 2000US-0224519P.
 PR 14-AUG-2000; 2000US-0225213P.
 PR 14-AUG-2000; 2000US-0225214P.
 PR 14-AUG-2000; 2000US-0225266P.
 PR 14-AUG-2000; 2000US-0225267P.
 PR 14-AUG-2000; 2000US-0225268P.
 PR 14-AUG-2000; 2000US-0225270P.
 PR 14-AUG-2000; 2000US-0225447P.
 PR 14-AUG-2000; 2000US-0225757P.
 PR 14-AUG-2000; 2000US-0225758P.
 PR 14-AUG-2000; 2000US-0225759P.
 PR 18-AUG-2000; 2000US-0226279P.
 PR 22-AUG-2000; 2000US-0226681P.
 PR 22-AUG-2000; 2000US-0226686P.
 PR 22-AUG-2000; 2000US-0227182P.
 PR 23-AUG-2000; 2000US-0227009P.
 PR 30-AUG-2000; 2000US-0228924P.
 PR 01-SEP-2000; 2000US-0229287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.
 PR 01-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 05-SEP-2000; 2000US-0229513P.
 PR 06-SEP-2000; 2000US-0230437P.
 PR 06-SEP-2000; 2000US-0230438P.
 PR 08-SEP-2000; 2000US-0231242P.
 PR 08-SEP-2000; 2000US-0231243P.
 PR 08-SEP-2000; 2000US-0231244P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 08-SEP-2000; 2000US-0231414P.
 PR 08-SEP-2000; 2000US-0232080P.
 PR 08-SEP-2000; 2000US-0232081P.

PR 12-SEP-2000; 2000US-0231368P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 14-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 13-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239315P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.

PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
(HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-488783/53.
XX N-PSDB; AAG26053.
PT New nucleic acid molecules encoding 461 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives.
XX Claim 11; SEQ ID NO 1019; 980pp; English.
XX The invention relates to isolated nucleic acid molecules and their
CC encoded secreted proteins. The nucleic acids and proteins are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC in diagnosing a pathological condition or susceptibility to a
CC pathological condition. Antibodies to the proteins can also be used in
CC alleviating symptoms associated with the disorders and in diagnostic
CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays
CC (ELISA). Disorders which are diagnosed or treated include autoimmune
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
CC nervous system disorders e.g. Alzheimer's disease, infections caused by
CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
CC and many other disorders listed in the specification. The polypeptides
CC can also be used to aid wound healing and epithelial cell proliferation,
CC to prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. The present
CC sequence represents a novel secreted protein of the invention. Note: The
CC sequence data for this patent did not form part of the printed

Query Match 87.8%; Score 43; DB 4; Length 429;
Best Local Similarity 77.8%; Pred. No. 2.7e+02;

Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGSPGSPGQ 9
Db 61 AGSPGAPGQ 69
:|||||:

RESULT 6

ABU55135
ID ABU55135 standard; protein; 429 AA.

XX ABU55135;

XX 18-MAR-2003 (first entry)

DE Human novel polypeptide #222.

XX Human; neural disorder; immune system disorder; renal disorder;

KW muscular disorder; respiratory disease; reproductive disorder;

KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;

KW hyperproliferative disorder; inflammatory disease; allergic reaction;
KW blood related disorder; cancer; immunosuppressive; antiinflammatory;
KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
KW haemostatic; antiarteriosclerotic.

XX Homo sapiens.

XX US2002132753-A1.

XX 19-SEP-2002.

XX 17-JAN-2001; 2001US-00764864.

XX 31-JAN-2000; 2000US-0179065P.

XX 04-FEB-2000; 2000US-0180628P.

XX 28-JUN-2000; 2000US-0214886P.

XX 07-JUL-2000; 2000US-0216647P.

XX 07-JUL-2000; 2000US-0216880P.

XX 11-JUL-2000; 2000US-0217487P.

XX 11-JUL-2000; 2000US-0217496P.

XX 14-JUL-2000; 2000US-0218290P.

XX 26-JUL-2000; 2000US-0220963P.

XX 26-JUL-2000; 2000US-0220964P.

XX 14-AUG-2000; 2000US-0224518P.

XX 14-AUG-2000; 2000US-0224519P.

XX 14-AUG-2000; 2000US-0225267P.

XX 14-AUG-2000; 2000US-0225270P.

XX 14-AUG-2000; 2000US-0225447P.

XX 14-AUG-2000; 2000US-0225757P.

XX 14-AUG-2000; 2000US-0225758P.

XX 22-AUG-2000; 2000US-0226868P.

XX 30-AUG-2000; 2000US-0228924P.

XX 01-SEP-2000; 2000US-0229287P.

XX 01-SEP-2000; 2000US-0229343P.

XX 01-SEP-2000; 2000US-0229344P.

XX 01-SEP-2000; 2000US-0229345P.

XX 05-SEP-2000; 2000US-0229509P.

XX 05-SEP-2000; 2000US-0229513P.

XX 08-SEP-2000; 2000US-0231413P.

XX 21-SEP-2000; 2000US-0234223P.

XX 21-SEP-2000; 2000US-0234274P.

XX 25-SEP-2000; 2000US-0234997P.

XX 27-SEP-2000; 2000US-0235834P.

XX 29-SEP-2000; 2000US-0236327P.

XX 29-SEP-2000; 2000US-0236367P.

XX 29-SEP-2000; 2000US-0236368P.

XX 29-SEP-2000; 2000US-0236369P.

XX 29-SEP-2000; 2000US-0236370P.

XX 02-OCT-2000; 2000US-0236802P.

XX 02-OCT-2000; 2000US-0237037P.

XX 02-OCT-2000; 2000US-0237038P.

XX 02-OCT-2000; 2000US-0237039P.

XX 13-OCT-2000; 2000US-0237040P.

XX 20-OCT-2000; 2000US-0239335P.

XX 20-OCT-2000; 2000US-0240960P.

XX 20-OCT-2000; 2000US-0241785P.

XX 01-NOV-2000; 2000US-0241809P.

XX 17-NOV-2000; 2000US-0244617P.

XX 08-DEC-2000; 2000US-0249299P.

XX 08-DEC-2000; 2000US-0251856P.

XX 08-DEC-2000; 2000US-0251868P.

XX 08-DEC-2000; 2000US-0251869P.

XX (ROSE/) ROSEN C A.

XX (RUBE/) RUBEN S M.

XX (BARA/) BARASH S C.

XX Rosen CA, Ruben SM, Barash SC;

XX WIPI: 2003-147444/14.

XX N-PSDB; ABX73394.

XX

PT New polypeptides and nucleic acids, useful in gene therapy for treating,
PT inhibiting or preventing e.g. neural, immune system, muscular,
PT respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
PT renal disorders.

XX Claim 11; SEQ ID NO 1019; 402pp; English.

XX The invention relates to human novel polypeptides and their associated
XX polynucleotides. The polypeptides and polynucleotides are useful in gene
XX therapy for treating, inhibiting or preventing neural disorders, immune
XX system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
XX and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
XX nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
XX gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
XX (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
XX heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
XX renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
XX leukaemia), inflammatory diseases (e.g. septic shock, bursitis and
XX appendicitis), allergic reactions and conditions (e.g. asthma), blood
XX related disorders (e.g. thrombosis, atherosclerosis and myocardial
XX infarction) and cancerous diseases. Sequences ABU54914-ABU55699 and
XX CC ABU55748 represent human novel polypeptides of the invention

XX Sequence 429 AA;

XX Query Match 87.8%; Score 43; DB 6; Length 429;
XX Best Local Similarity 77.8%; Pred. NO. 2.7e+02;
XX Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGSPGSPGQ 9
Db 61 AGSPGAPGQ 69

RESULT 7
ABU52683
ID ABU52683 standard; protein; 957 AA.

XX ABU52683;
XX 14-APR-2003 (first entry)

XX Human cell structure and mobility-associated protein from DKF2phtb2_2b5.
XX Human; gene therapy; vaccine; disease treatment; detection.

XX Homo sapiens.
XX WO200112659-A2.

XX 22-FEB-2001.
XX 18-AUG-2000; 2000WO-IB001496.

XX 18-AUG-1999; 99US-0149499P.
XX 28-SEP-1999; 99US-0156503P.

XX (GEHU-) GERMAN HUMAN GENOME PROJECT.
XX Wiemann S;

XX WIPI: 2001-327840/34.
XX N-PSDB; ABX71234.

XX Nucleic acids having the sequences of clones isolated from libraries of
XX different human tissues, useful in recombinant DNA methodologies.

XX Claim 21; Page 202-203; 1095pp; English.

XX This invention describes novel polynucleotides and polypeptides isolated
XX from human cDNA libraries which can be used for gene therapy or in
XX vaccines. The polynucleotides of the invention and antibodies encoded by
XX them may be used in the prevention, diagnosis and treatment of diseases

CC associated with inappropriate polypeptide expression. The products of the
 CC invention may also be used to identify modulators of expression and
 CC activity and to down regulate expression and activity. The antibodies of
 CC the invention may also be used as diagnostic agents for detecting the
 CC presence of polypeptides in samples. This sequence represents a
 CC polypeptide described in the disclosure of the invention
 XX
 SQ Sequence 957 AA;

Query Match 87.8%; Score 43; DB 4; Length 957;
 Best Local Similarity 77.8%; Pred. No. 5.9e+02;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 SGSPGSPGQ 9
 :|||||:
 Db 589 AGSPGAPGQ 597

RESULT 8
 ABR40002
 ID ABR40002 standard; protein; 1606 AA.

XX AC ABR40002;
 XX DT 03-JUN-2003 (first entry)
 XX DE Human collagen XXII alternatively spliced variant protein.

XX Human; collagen XXII; vulnery; antiarthritic; osteopathic;
 KW ophthalmological; gene therapy; connective tissue disorder; joint;
 KW tendon; cartilage; basement membrane; wound repair; osteoarthritis;
 KW osteogenesis imperfecta; dystrophic epidermolysis bullosa; polymyositis;
 KW heart valve disorder; heart disorder; eye disorder; hair loss.

XX OS Homo sapiens.

XX Key Location/Qualifiers
 FT Misc-difference 472. .473
 FT /note= "Encoded by AAGACC"

XX WO2003012121-A2.

XX PD 13-FEB-2003.

XX PF 22-JUL-2002; 2002WO-US023075.

XX PR 31-JUL-2001; 2001US-0309158P.

XX PA (GEO) GEN HOSPITAL CORP.

XX PI Burgeson RE, Koch M, Bruckner-Tuderman L, Keene DR, Brunken WJ;

XX WPI; 2003-239527/23.

XX DR N-PSDB; ACC00270.

XX New isolated collagen XXII nucleic acids and polypeptides, useful for
 PT diagnosing or treating conditions related to aberrant activity or
 PT expression of collagen XXII, e.g. connective tissue disorders, or for
 PT promoting wound repair.

XX Example 1; Page; 118pp; English.

XX The present sequence is the protein sequence for an alternatively spliced
 CC variant of human collagen XXII. This sequence lacks residues 1005-1024 of
 CC ABR40001. Collagen XXII and its coding sequence are useful for diagnosing
 CC or treating conditions related to aberrant activity or expression of the
 CC collagen XXII, such as conditions involving aberrant or deficient tissue
 CC strength, e.g. connective tissue disorder, disorders of the joint,
 CC tendon, cartilage or basement membrane, or for promoting wound repair.
 CC Collagen XXII is useful as a reagent or target in assays for the
 CC treatment or diagnosis of collagen XXII-mediated or -related disorders,
 CC e.g. osteoarthritis, osteogenesis imperfecta, dystrophic epidermolysis
 CC bullosa, polymyositis, disorders of the heart valve or other regions of

CC the heart, disorders of the eyes, hair related conditions such as hair
 CC loss. Note: The present sequence was not shown in the specification but
 CC was derived from information given
 XX

SQ Sequence 1606 AA;

Query Match 87.8%; Score 43; DB 6; Length 1606;
 Best Local Similarity 77.8%; Pred. No. 9.7e+02;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 SGSPGSPGQ 9
 :|||||:
 Db 1309 NGSPGSPGE 1317

RESULT 9
 ABR40001
 ID ABR40001 standard; protein; 1626 AA.

XX AC ABR40001;

XX DT 03-JUN-2003 (first entry)

XX DE Human collagen XXII.

XX Human; collagen XXII; vulnery; antiarthritic; osteopathic;
 KW ophthalmological; gene therapy; connective tissue disorder; joint;
 KW tendon; cartilage; basement membrane; wound repair; osteoarthritis;
 KW osteogenesis imperfecta; dystrophic epidermolysis bullosa; polymyositis;
 KW heart valve disorder; heart disorder; eye disorder; hair loss.

XX OS Homo sapiens.

XX Key Location/Qualifiers
 FT Misc-difference 472. .473
 FT /note= "Encoded by AAGACC"

XX WO2003012121-A2.

XX PD 13-FEB-2003.

XX PF 22-JUL-2002; 2002WO-US023075.

XX PR 31-JUL-2001; 2001US-0309158P.

XX PA (GEO) GEN HOSPITAL CORP.

XX PI Burgeson RE, Koch M, Bruckner-Tuderman L, Keene DR, Brunken WJ;

XX WPI; 2003-239527/23.

XX DR N-PSDB; ACC00269.

XX New isolated collagen XXII nucleic acids and polypeptides, useful for
 PT diagnosing or treating conditions related to aberrant activity or
 PT expression of collagen XXII, e.g. connective tissue disorders, or for
 PT promoting wound repair.

XX Claim 10; Fig 2; 118pp; English.

XX The present sequence is the protein sequence for human collagen XXII.
 CC Collagen XXII and its coding sequence are useful for diagnosing or
 CC treating conditions related to aberrant activity or expression of the
 CC collagen XXII, such as conditions involving aberrant or deficient tissue
 CC strength, e.g. connective tissue disorder, disorders of the joint,
 CC tendon, cartilage or basement membrane, or for promoting wound repair.
 CC Collagen XXII is useful as a reagent or target in assays for the
 CC treatment or diagnosis of collagen XXII-mediated or -related disorders,
 CC e.g. osteoarthritis, osteogenesis imperfecta, dystrophic epidermolysis
 CC bullosa, polymyositis, disorders of the heart valve or other regions of
 CC the heart, disorders of the eyes, hair related conditions such as hair
 CC loss

XX Sequence 1626 AA;

SQ

Query Match 87.8%; Score 43; DB 6; Length 1626;
 Best Local Similarity 77.8%; Pred. No. 9.9e+02;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGSPGSPGQ 9
 DB 1329 NGSFGSPGE 1337

RESULT 10
 AAB06022
 ID AAB06022 standard; protein; 104 AA.
 XX AAB06022;
 AC
 XX 23-OCT-2000 (first entry)
 DT
 XX Polar gelatin P monomer.
 DE
 XX Polar gelatin; P monomer; P tetramer; P4; synthetic; collagen-like;
 KW photographic paper; triacetate cellulose film; Pichia pastoris;
 KW tubular silver halide emulsion; fed-batch fermentation.
 XX
 OS Synthetic.
 XX
 XX EF1014176-A2.
 PN
 XX 28-JUN-2000.
 PD
 XX 17-DEC-1999; 99BP-00204382.
 PF
 XX 23-DEC-1998; 98US-00219849.
 PR
 XX (FUJF) FUJI PHOTO FILM BV.
 PA
 XX De Wolf A, Werten MWT, Wisselink HW, Jansen-Van Den Bosch TJ;
 PI Toda Y, Van Heerde GV, Bouwstra JB;
 PI WPI; 2000-402280/35.
 DR
 XX Tabular silver halide emulsion used in the manufacture of photographic
 PT paper and triacetate cellulose film.
 PT
 XX Example 1; Page 7; 12pp; English.
 PS
 XX The present sequence is P monomer, a polar gelatin encoded by a synthetic
 CC gene constructed by overlap extension PCR. The gene was designed to have
 CC the codon usage of Pichia pastoris highly expressed genes. A nucleic acid
 CC encoding a P tetramer (P4) was constructed using the P monomer PCR
 CC product. The P monomer and tetramer fragments were cloned into vector
 CC pPIC19, which was linearised and used to transform Pichia pastoris strain
 CC GS115. Fed-batch fermentation resulted in the production of a non-natural
 CC gelatin-like protein which may be used in the production of tabular
 CC silver halide emulsions. These emulsions are used in the manufacture of
 CC photographic paper and triacetate cellulose film. The invention allows
 CC the production of large amounts of collagen-like polypeptides without
 CC requiring expensive media, expensive expression hosts or non-secreting
 CC expression hosts
 XX
 SQ Sequence 104 AA;

Query Match 85.7%; Score 42; DB 3; Length 104;
 Best Local Similarity 87.5%; Pred. No. 98;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GSPGSPGQ 9
 DB 22 GSPGNPGQ 29

RESULT 11
 ABG31593

Query Match 85.7%; Score 42; DB 5; Length 161;
 Best Local Similarity 87.5%; Pred. No. 1.5e+02;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GSPGSPGQ 9
 DB 22 GSPGNPGQ 29

RESULT 12
 ABB98499
 ID ABB98499 standard; protein; 161 AA.
 XX ABB98499;
 AC
 XX 15-NOV-2002 (first entry)
 DT
 XX Mature H-gelatin.
 DE

ID ABG31593 standard; protein; 161 AA.
 XX
 AC ABG31593;
 DT
 XX 05-NOV-2002 (first entry)
 DT
 XX Synthetic mature hydrophilic gelatin (H-gelatin) protein.
 DE
 XX Silver-dispersing; collagen; nucleation peptiser; growth peptiser;
 KW silver-binding; photographic manufacturing process; major face;
 KW thin tabular silver halide crystal; photographic emulsion layer;
 KW light-sensitive material; hydrophilic gelatin; H-gelatin.
 XX
 OS Synthetic.
 XX
 XX WO200252342-A1.
 PN
 XX 04-JUL-2002.
 PD
 XX 21-DEC-2001; 2001WO-NL000933.
 PF
 XX 27-DEC-2000; 2000EP-00204775.
 PR
 XX (FUJF) FUJI PHOTO FILM BV.
 PA
 XX Bouwstra JB, Boegels G, Toda Y;
 PI WPI; 2002-599583/64.
 DR
 XX Novel recombinant collagen-like polypeptide for photographic manufacture,
 PT has various region(s) having amino acids, region having silver-binding
 PT amino acids and the polypeptide does not have methionine or cysteine
 PT residues.
 XX
 XX Example 1; Page 15; 25pp; English.
 PS
 XX The present invention relates to a new recombinant collagen-like
 CC polypeptide suitable as a nucleation or growth peptiser. The polypeptide
 CC of the invention comprises region(s) A(n) or A(g) having a maximum of 100
 CC amino acids, where A(n) is delimited by the extreme two silver-binding
 CC histidines and A(g) consisting silver-binding amino acids and a region
 CC B(n) or B(g), having 50 amino acids and the polypeptide not containing
 CC methionine or cysteine. The invention is useful as peptiser in
 CC photographic manufacturing processes and is also useful to manufacture
 CC thin tabular silver halide crystals having major faces. The invention is
 CC further useful in photographic emulsion layers of a light-sensitive
 CC material. The polypeptide of the invention provides photographic emulsion
 CC with an increased number of tabular crystals and also photographic
 CC emulsions with good dispersion stability of the silver halide crystals.
 CC The polypeptide provides steric repulsion without shielding of the
 CC crystal surface. The present amino acid sequence represents the
 CC hydrophilic gelatin (H-gelatin) protein as described in the invention
 XX
 SQ Sequence 161 AA;

```

XX Gelatin; plasma expander; plasma substitute.
KW Pichia pastoris.
OS Synthetic.
XX EP1238675-A1.
XX 11-SEP-2002.
XX
XX 06-MAR-2001; 2001EP-00200837.
XX
XX 06-MAR-2001; 2001EP-00200837.
XX (FUJF ) FUJI PHOTO FILM BV.
XX Bouwstra JB, Toda Y;
PI WPI; 2002-645919/70.
XX
XX Use of recombinant gelatin-like protein which is essentially free of
PT hydroxyproline as a plasma expander.
XX
XX Example 1; Page 9; 14pp; English.
XX
XX The present invention relates to a recombinant gelatin-like protein,
CC which is in essence free of hydroxyproline. The gelatin-like protein is
CC useful as a plasma expander and a plasma substitute. The absence of
CC hydroxyproline prevents compositions comprising the gelatin-like protein
CC from gelling and thus allows the use of high molecular weight proteins in
CC order to establish a suitable colloid osmotic pressure. PCR primers
CC ABQ0666-ABQ0673 were used to clone a synthetic gene encoding a
CC hydrophilic gelatin with six histidine residues (H-gelatin). The
CC synthetic gene was designed to have the codon usage of Pichia pastoris
CC highly expressed genes. The resulting PCR product of 0.3 kb was inserted
CC in cloning vector pMTL23 to generate pMTL23-P. PCR primers ABQ0674-
CC ABQ0677 were used in a separate PCR assay and the resulting PCR product
CC of 0.18 kb was cloned into pMTL23-P. The resulting vector was then cut
CC with EcoRI/XhoI, after which the insert was cloned into EcoRI/XhoI
CC digested P. pastoris expression vector pPIC9, to yield vector pPIC9-H1,
CC which encodes the mature (processed) H-gelatin protein (AB898499). The
CC mature H-gelatin protein has a molecular weight of 15.1 kDa and an
CC isoelectric point of 5.1
XX
XX Sequence 161 AA;
SQ
Query Match 85.7%; Score 42; DB 5; Length 161;
Best Local Similarity 87.5%; Pred. No. 1.5e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 GSPGSPGQ 9
Db 22 GSPGNPQQ 29
|||||
|||||
RESULT 13
ID ABB67768
XX ABB67768 standard; protein; 317 AA.
XX
XX ABB67768;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 30096.
DE
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
KW Drosophila melanogaster.
OS
XX WO200171042-A2.
PN
XX 27-SEP-2001.
PD
Gelatin; plasma expander; plasma substitute.
KW Pichia pastoris.
OS Synthetic.
XX EP1238675-A1.
XX 11-SEP-2002.
XX
XX 06-MAR-2001; 2001EP-00200837.
XX
XX 06-MAR-2001; 2001EP-00200837.
XX (FUJF ) FUJI PHOTO FILM BV.
XX Bouwstra JB, Toda Y;
PI WPI; 2002-645919/70.
XX
XX Use of recombinant gelatin-like protein which is essentially free of
PT hydroxyproline as a plasma expander.
XX
XX Example 1; Page 9; 14pp; English.
XX
XX The present invention relates to a recombinant gelatin-like protein,
CC which is in essence free of hydroxyproline. The gelatin-like protein is
CC useful as a plasma expander and a plasma substitute. The absence of
CC hydroxyproline prevents compositions comprising the gelatin-like protein
CC from gelling and thus allows the use of high molecular weight proteins in
CC order to establish a suitable colloid osmotic pressure. PCR primers
CC ABQ0666-ABQ0673 were used to clone a synthetic gene encoding a
CC hydrophilic gelatin with six histidine residues (H-gelatin). The
CC synthetic gene was designed to have the codon usage of Pichia pastoris
CC highly expressed genes. The resulting PCR product of 0.3 kb was inserted
CC in cloning vector pMTL23 to generate pMTL23-P. PCR primers ABQ0674-
CC ABQ0677 were used in a separate PCR assay and the resulting PCR product
CC of 0.18 kb was cloned into pMTL23-P. The resulting vector was then cut
CC with EcoRI/XhoI, after which the insert was cloned into EcoRI/XhoI
CC digested P. pastoris expression vector pPIC9, to yield vector pPIC9-H1,
CC which encodes the mature (processed) H-gelatin protein (AB898499). The
CC mature H-gelatin protein has a molecular weight of 15.1 kDa and an
CC isoelectric point of 5.1
XX
XX Sequence 161 AA;
SQ
Query Match 85.7%; Score 42; DB 5; Length 161;
Best Local Similarity 87.5%; Pred. No. 1.5e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 GSPGSPGQ 9
Db 22 GSPGNPQQ 29
|||||
|||||
RESULT 13
ID ABB67768
XX ABB67768 standard; protein; 317 AA.
XX
XX ABB67768;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 30096.
DE
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
KW Drosophila melanogaster.
OS
XX WO200171042-A2.
PN
XX 27-SEP-2001.
PD
Gelatin; plasma expander; plasma substitute.
KW Pichia pastoris.
OS Synthetic.
XX EP1238675-A1.
XX 11-SEP-2002.
XX
XX 06-MAR-2001; 2001EP-00200837.
XX
XX 06-MAR-2001; 2001EP-00200837.
XX (FUJF ) FUJI PHOTO FILM BV.
XX Bouwstra JB, Toda Y;
PI WPI; 2002-645919/70.
XX
XX Use of recombinant gelatin-like protein which is essentially free of
PT hydroxyproline as a plasma expander.
XX
XX Example 1; Page 9; 14pp; English.
XX
XX The present invention relates to a recombinant gelatin-like protein,
CC which is in essence free of hydroxyproline. The gelatin-like protein is
CC useful as a plasma expander and a plasma substitute. The absence of
CC hydroxyproline prevents compositions comprising the gelatin-like protein
CC from gelling and thus allows the use of high molecular weight proteins in
CC order to establish a suitable colloid osmotic pressure. PCR primers
CC ABQ0666-ABQ0673 were used to clone a synthetic gene encoding a
CC hydrophilic gelatin with six histidine residues (H-gelatin). The
CC synthetic gene was designed to have the codon usage of Pichia pastoris
CC highly expressed genes. The resulting PCR product of 0.3 kb was inserted
CC in cloning vector pMTL23 to generate pMTL23-P. PCR primers ABQ0674-
CC ABQ0677 were used in a separate PCR assay and the resulting PCR product
CC of 0.18 kb was cloned into pMTL23-P. The resulting vector was then cut
CC with EcoRI/XhoI, after which the insert was cloned into EcoRI/XhoI
CC digested P. pastoris expression vector pPIC9, to yield vector pPIC9-H1,
CC which encodes the mature (processed) H-gelatin protein (AB898499). The
CC mature H-gelatin protein has a molecular weight of 15.1 kDa and an
CC isoelectric point of 5.1
XX
XX Sequence 161 AA;
SQ
Query Match 85.7%; Score 42; DB 4; Length 317;
Best Local Similarity 77.8%; Pred. No. 2.9e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 SGSPGSPGQ 9
Db 48 TGSPGPGQ 56
|||||
|||||
RESULT 14
ID AAB06023
XX AAB06023 standard; protein; 401 AA.
XX
XX AAB06023;
AC
XX
XX 23-OCT-2000 (first entry)
DT
XX
XX Polar gelatin P tetramer, P4.
DE
XX
XX Polar gelatin; P monomer; P tetramer; P4; synthetic; collagen-like;
KW photographic paper; triacetate cellulose film; Pichia pastoris;
KW tabular silver halide emulsion; fed-batch fermentation.
XX
XX Synthetic.
OS
XX EP1014176-A2.
PN
XX
XX 28-JUN-2000.
PD
XX
XX 17-DEC-1999; 99EP-00204382.
PF
XX
XX 23-DEC-1998; 98US-00219849.
PR
XX (FUJF ) FUJI PHOTO FILM BV.
PA
XX De Wolf A, Werten MWT, Wisselink HW, Janssen-Van Den Bosch TJ;
PI Toda Y, Van Heerde GV, Bouwstra JB;
XX WPI; 2000-402280/35.
DR
XX Tabular silver halide emulsion used in the manufacture of photographic
PT

```


PT paper and triacetate cellulose film.
 XX
 PS Example 1; Page 7; 12pp; English.
 XX

CC The present sequence is P tetramer (P4). A nucleic acid encoding P4 was
 CC constructed using a synthetic gene encoding P monomer, a non-natural
 CC polar gelatin. The P monomer gene was designed to have the codon usage of
 CC Pichia pastoris highly expressed genes. The P monomer and tetramer
 CC fragments were cloned into vector pPIC19, which was linearised and used
 CC to transform Pichia pastoris strain GS115. Fed-batch fermentation
 CC resulted in the production of a non-natural gelatin-like protein which
 CC emulsions are used in the manufacture of photographic paper and
 CC triacetate cellulose film. The invention allows the production of large
 CC amounts of collagen-like polypeptides without requiring expensive media,
 CC expensive expression hosts or non-secreting expression hosts
 XX
 SQ Sequence 401 AA;

Query Match 85.7%; Score 42; DB 3; Length 401;
 Best Local Similarity 87.5%; Pred. No. 3.6e+02;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GSPGSPGQ 9
 ||||:||||
 Db 22 GSPGNFGQ 29

RESULT 15
 AAY72375
 ID AAY72375 standard; protein; 599 AA.
 XX
 AC AAY72375;
 XX
 DT 24-APR-2001 (first entry)
 XX
 DE Amphiphilic recombinant collagen-like polymer, N1N2P4.
 XX
 KW Recombinant protein; collagen-like polymer; gelatin-like polymer; N1N2P4;
 KW stabiliser; oil-in-water emulsion; foodstuff; pharmaceutical product;
 KW cosmetic product; photography.
 XX
 OS Synthetic.
 XX
 PN EP1063565-A1.
 XX
 PD 27-DEC-2000.
 XX
 PF 23-JUN-2000; 2000EP-00202206.
 XX
 PR 24-JUN-1999; 99EP-00202047.
 XX
 PA (FUJIF) FUJI PHOTO FILM BV.
 XX
 PI Olijve JH, Bouwstra JB, De Wolf FA, Werten MWT, Wisselink HW;
 PI Wind RD, Van Den Bosch TJ, Toda Y;
 XX
 DR WPI; 2001-125578/14.
 XX
 PT Oil-in-water emulsions for preparing foodstuffs, pharmaceutical product
 PT or cosmetic product, comprises recombinant collagen-like polymer as
 PT stabilizer.
 XX
 PS Example 6; Page 24-26; 31pp; English.
 XX

CC The present sequence is a synthetic amphiphilic recombinant collagen-
 CC like (or gelatin-like) polymer, N1N2P4 which contains two different non-
 CC polar modules (N1) and four polar modules (P4). The N2 module is similar
 CC to the N1 module, but differs mainly in the presence of a cluster of
 CC methionine and charged residues at its C-terminal side. This polymer
 CC exhibits an amphiphilic structure and is used as a stabiliser for oil-in
 CC water emulsions. The oil-in-water emulsion is used for producing
 CC foodstuffs, pharmaceutical products or cosmetic products by combining

CC with nutritionally, pharmaceutically, and cosmetically suitable
 CC ingredients. It is also useful in photography
 XX
 SQ Sequence 599 AA;
 Query Match 85.7%; Score 42; DB 4; Length 599;
 Best Local Similarity 87.5%; Pred. No. 5.3e+02;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GSPGSPGQ 9
 ||||:||||
 Db 220 GSPGNFGQ 227

Search completed: July 18, 2005, 13:40:29
 Job time : 77.78 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2005, 13:09:46 ; Search time 20.02 Seconds

(without alignments)
62.478 Million cell updates/sec

Title: SEQ4

Perfect score: 67

Sequence: 1 satttgsspgptq 13

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	63	94.0	513	1 EUTQ1	cellulose 1,4-beta
2	63	94.0	513	2 S45380	cellulose 1,4-beta
3	59	88.1	513	2 S11439	cellulose 1,4-beta
4	55	82.1	302	2 S71334	acetyl xylan ester
5	44	65.7	615	2 B39897	GTPase-activating
6	44	65.7	663	2 A39897	GTPase-activating
7	42	62.7	519	2 S78089	G-protein signalin
8	42	62.7	800	2 T19627	hypothetical prote
9	42	62.7	1270	2 T09194	adaptor protein in
10	41	61.2	482	2 A30198	dihydrolipamide S
11	41	61.2	622	2 T27155	hypothetical prote
12	41	61.2	1777	2 T34369	hypothetical prote
13	40	59.7	173	2 S16162	crucipain (EC 3.4.
14	40	59.7	186	2 T31347	hypothetical prote
15	40	59.7	204	2 T02386	hypothetical prote
16	40	59.7	305	2 S41860	gene Nkx-1.1 prote
17	40	59.7	467	2 A60867	cysteine proteinas
18	40	59.7	467	2 A45629	cysteine proteinas
19	40	59.7	525	2 T44445	chitinase (EC 3.2.
20	40	59.7	841	2 B71212	hypothetical prote
21	40	59.7	3020	2 A43932	mucin 2 precursor
22	39.5	59.0	88	2 A38085	S-layer glycoprote
23	39.5	59.0	827	2 A37849	S-layer protein -
24	39	58.2	108	2 F72469	hypothetical prote
25	39	58.2	126	2 I61260	synapsin II - mous
26	39	58.2	324	2 S49586	cysteine synthase
27	39	58.2	382	2 T49762	hypothetical prote
28	39	58.2	417	2 JC7092	Faul protein - fis
29	39	58.2	479	2 D30411	synapsin Iib - rat

30	39	58.2	513	2 F86320	hypothetical prote
31	39	58.2	556	2 S06838	gamma-aminobutyric
32	39	58.2	559	2 B56731	chromatin assembly
33	39	58.2	586	2 C30411	synapsin Iia - rat
34	39	58.2	843	2 S33442	EF protein - Strept
35	39	58.2	861	2 S77409	hypothetical prote
36	39	58.2	1379	2 S64603	YTA7 protein - yea
37	39	58.2	1822	2 S33441	EF protein - Strept
38	38	56.7	119	2 T72487	hypothetical prote
39	38	56.7	303	2 T02588	hypothetical prote
40	38	56.7	312	2 D87475	rare lipoprotein A
41	38	56.7	340	2 H81742	major outer membra
42	38	56.7	365	2 T34759	oligopeptide ABC t
43	38	56.7	391	1 JQ1626	attachment protein
44	38	56.7	494	2 A88474	protein C05D10.1 [
45	38	56.7	540	2 S21825	vicilin-like stora

ALIGNMENTS

RESULT 1

EUTQ1

cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) I precursor - fungus (Trichoderma reesei)

N;Alternate_names: exo-cellobiohydrolase I

C;Species: Trichoderma reesei

C;Date: 03-Aug-1984 #sequence_revision 03-Aug-1984 #text_change 09-Jul-2004

C;Accession: A00902

R;Shoemaker, S.; Schweickart, V.; Ladner, M.; Gelfand, D.; Kwok, S.; Myambo, K.; Innis, N

Bio/Technology 1, 691-696, 1983

A;Title: Molecular cloning of exo-cellobiohydrolase I derived from Trichoderma reesei str

A;Reference number: A00902

A;Accession: A00902

A;Molecule type: DNA

A;Residues: 1-513 <SHO>

A;Cross-references: UNIPROT:P00725

A;Experimental source: Strain L27

C;Comment: This is the most abundantly produced cellulase in this filamentous fungus; it

C;Genetics:

A;Gene: CBH1

A;Introns: 154/2; 386/3

C;Function:

A;Description: catalyzes the hydrolysis of 1,4-beta-D-glucosidic bonds in cellulose to re

C;Superfamily: cellulose 1,4-beta-cellobiosidase I; fungal cellulose-binding domain homol

C;Keywords: glycosidase; hydrolase; polysaccharide degradation

F;1-17/Domain: signal sequence #status predicted <SIG>

F;18-513/Product: cellulose 1,4-beta-cellobiosidase I #status predicted <WAT>

F;482-513/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 94.0%; Score 63; DB 1; Length 513;

Best Local Similarity 100.0%; Pred. No. 0.025;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 ATTGSSPGPTQ 13

Db 468 ATTGSSPGPTQ 479

RESULT 2

S45380

cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) - fungus (Trichoderma koningii)

C;Species: Trichoderma koningii

C;Date: 20-Oct-1994 #sequence_revision 10-Nov-1995 #text_change 22-Jun-1999

C;Accession: S45380

R;Wey, T.T.; Hseu, T.H.; Huang, L.

Curr. Microbiol. 28, 31-39, 1994

A;Title: Molecular cloning and sequence analysis of the cellobiohydrolase I gene from Tr

A;Reference number: S45380; MUID:94100788; PMID:7764306

A;Accession: S45380

A;Molecule type: DNA

A;Residues: 1-513 <WEY>

A;Cross-references: EMBL:X69976; NID:9457422; PIDN:CAA49596.1; PID:9457423

C;Genetics:

A;Introns: 154/2; 386/3
 C;Superfamily: cellulose 1,4-beta-cellobiosidase I; fungal cellulose-binding domain homolog
 C;Keywords: glycosidase; hydrolase; polysaccharide degradation
 F;482-513/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 94.0%; Score 63; DB 2; Length 513;
 Best Local Similarity 100.0%; Pred. No. 0.025;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ATTTGSSPGPTQ 13
 |||||
 Db 468 ATTTGSSPGPTQ 479

RESULT 3
 S11439
 cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) - fungus (*Trichoderma viride*)
 C;Species: *Trichoderma viride*
 C;Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004
 C;Accession: S11439
 R;Cheng, C.; Tsukagoshi, N.; Uchida, S.
 Nucleic Acids Res 18, 5559, 1990
 A;Title: Nucleotide sequence of the cellobiohydrolase gene from *Trichoderma viride*.
 A;Reference number: S11439; MUID:91016856; PMID:2216737
 A;Accession: S11439
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-513 <CHE>
 A;Cross-references: UNIPROT:P19355; EMBL:X53931; NID:G5196; PIDN:CAA37878.1; PID:G295937
 C;Genetics:

A;Introns: 154/2; 386/3
 C;Superfamily: cellulose 1,4-beta-cellobiosidase I; fungal cellulose-binding domain homolog
 C;Keywords: glycosidase; hydrolase; polysaccharide degradation
 F;482-513/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 88.1%; Score 59; DB 2; Length 513;
 Best Local Similarity 91.7%; Pred. No. 0.11;
 Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 ATTTGSSPGPTQ 13
 |||||
 Db 468 ATTTGSSPGPTQ 479

RESULT 4
 S71334
 acetyl xylan esterase precursor - fungus (*Trichoderma reesei*)
 C;Species: *Trichoderma reesei*
 C;Date: 23-Jul-1997 #sequence_revision 01-Aug-1997 #text_change 09-Jul-2004
 C;Accession: S71334
 R;Margolles-Clark, E.; Tenkanen, M.; Soederlund, H.; Penttilae, M.
 Eur. J. Biochem. 237, 553-560, 1996
 A;Title: Acetyl xylan esterase from *Trichoderma reesei* contains an active-site serine residue
 A;Reference number: S71334; MUID:96235218; PMID:8647098
 A;Accession: S71334
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-302 <VAR>
 A;Cross-references: UNIPROT:Q99034; EMBL:Z69256; NID:G1431619; PID:e220701; PID:G1431620
 C;Genetics:

F;1-20/Domain: signal sequence #status predicted <SIG>
 F;21-302/Product: acetyl xylan esterase #status predicted <WAT>
 F;271-302/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 82.1%; Score 55; DB 2; Length 302;
 Best Local Similarity 76.9%; Pred. No. 0.26;
 Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 SATTTGSSPGPTQ 13
 :||
 Db 256 TSRTGSSPGPTQ 268

RESULT 5

B39897
 GTPase-activating protein rap1GAP long form - human (fragment)

C;Species: *Homo sapiens* (man)
 C;Date: 24-Jan-1992 #sequence_revision 24-Jan-1992 #text_change 18-Jun-1993
 C;Accession: B39897

R;Rubinfeld, B.; Munemitsu, S.; Clark, R.; Conroy, L.; Watt, K.; Crosier, W.J.; McCormick
 Cell 65, 1033-1042, 1991
 A;Title: Molecular cloning of a GTPase activating protein specific for the Krev-1 protein

A;Reference number: A39897; MUID:91256304; PMID:1904317

A;Accession: B39897

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-615 <RUB>

A;Cross-references: GB:M64788

Query Match 65.7%; Score 44; DB 2; Length 615;

Best Local Similarity 61.5%; Pred. No. 29;

Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 SATTTGSSPGPTQ 13

|||

Db 570 STTGGSSPGPSR 582

RESULT 6

A39897
 GTPase-activating protein rap1GAP short form - human

C;Species: *Homo sapiens* (man)

C;Date: 24-Jan-1992 #sequence_revision 24-Jan-1992 #text_change 09-Jul-2004

C;Accession: A39897

R;Rubinfeld, B.; Munemitsu, S.; Clark, R.; Conroy, L.; Watt, K.; Crosier, W.J.; McCormick
 Cell 65, 1033-1042, 1991
 A;Title: Molecular cloning of a GTPase activating protein specific for the Krev-1 protein

A;Reference number: A39897; MUID:91256304; PMID:1904317

A;Accession: A39897

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-663 <RUB>

A;Cross-references: UNIPROT:P47736; GB:M64788; NID:G190855; PIDN:AAA60252.1; PID:G190856

Query Match 65.7%; Score 44; DB 2; Length 663;

Best Local Similarity 61.5%; Pred. No. 31;

Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 SATTTGSSPGPTQ 13

|||

Db 618 STTGGSSPGPSR 630

RESULT 7

S78089
 G-protein signaling regulator RGS3 - human

C;Species: *Homo sapiens* (man)

C;Date: 04-Dec-1997 #sequence_revision 12-Dec-1997 #text_change 09-Jul-2004

C;Accession: S78089; S68436

R;Druey, K.

submitted to the EMBL Data Library, May 1995

A;Reference number: S78089

A;Accession: S78089

A;Molecule type: mRNA

A;Residues: 1-519 <DRU>

A;Cross-references: UNIPROT:P49796; EMBL:U27655; NID:G1216368; PID:G1216369

R;Druey, K.M.; Blumber, K.J.; Kang, V.H.; Kehrl, J.H.

Nature 379, 742-746, 1996

A;Title: Inhibition of G-protein-mediated MAP kinase activation by a new mammalian gene
 A;Reference number: A58012; MUID:96178495; PMID:8602223

A;Accession: S68436

A;Status: nucleic acid sequence not shown

A;Molecule type: mRNA

A;Residues: 1-355; 'K', 357-519 <DRW>

A;Cross-references: EMBL:U27655

A;Experimental source: tonsil
A;Note: the sequence from Fig. 1b is inconsistent with that from Fig. 1a in having 356-1

Query Match 62.7%; Score 42; DB 2; Length 519;
Best Local Similarity 72.7%; Pred. No. 49;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 SATTGSSPCP 11
|||:||||
Db 67 SATSGSPGP 77

RESULT 8
T19627
hypothetical protein F36F2.5 - *Caenorhabditis elegans*
C;Species: *Caenorhabditis elegans*
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T19627; T21863
R;Kershaw, J.
submitted to the EMBL Data Library, April 1997
A;Reference number: Z19153
A;Accession: T19627
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-800 <WIL>
A;Cross-references: UNIPROT:O62237; EMBL:Z93778; PIDN:CAB07847.1; GSPDB:GN00019; CESP:F36F2
A;Experimental source: clone C31H5
R;Cottage, A.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19479
A;Accession: T21863
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-800 <W12>
A;Cross-references: EMBL:Z81532; PIDN:CAB04328.1; GSPDB:GN00019; CESP:F36F2.5
A;Experimental source: clone F36F2
C;Genetics:
A;Gene: CESP:F36F2.5
A;Map position: 1
A;Introns: 27/3; 51/2; 142/2; 191/3; 287/3; 333/3; 355/2; 405/1; 433/2; 484/1; 508/1; 56

Query Match 62.7%; Score 42; DB 2; Length 800;
Best Local Similarity 63.6%; Pred. No. 77;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 ATTTGSSPCP 12
|||:||||
Db 419 ATSTGNPAPT 429

RESULT 9
T09194
adaptor protein intersectin - African clawed frog
C;Species: *Xenopus laevis* (African clawed frog)
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
C;Accession: T09194
R;Yamabhai, M.; Hoffman, N.G.; Hardison, N.L.; McPherson, P.S.; Castagnoli, L.; Cesareni
J. Biol. Chem. 273, 31401-31407, 1998
A;Title: Intersectin, a novel adaptor protein with two eps15 homology and five src homol
A;Reference number: Z16605; MUID:99030416; PMID:9813051
A;Accession: T09194
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1270 <YAM>
A;Cross-references: UNIPROT:O42287; EMBL:AF032118; NID:g2642624; PIDN:AACT73068.1; PID:92
A;Experimental source: cell type oocyte
C;Function:
A;Description: involved in endocytosis
C;Keywords: endocytosis

Query Match 62.7%; Score 42; DB 2; Length 1270;
Best Local Similarity 80.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ATTTGSSPCP 11
|||:||||
Db 880 ATVTGSSPSP 889

RESULT 10
A30198
dihydrolipoamide S-acetyltransferase (EC 2.3.1.12) precursor - yeast (*Saccharomyces cerev*
N;Alternate names: protein N2374; protein YNL071W
C;Species: *Saccharomyces cerevisiae*
C;Date: 18-Oct-1989 #sequence_revision 18-Oct-1989 #text_change 09-Jul-2004
C;Accession: A30198; S53909; S63003; S63938
R;Niu, X.; Browning, K.S.; Behal, R.H.; Reed, L.J.
Proc. Natl. Acad. Sci. U.S.A. 85, 7546-7550, 1988
A;Title: Cloning and nucleotide sequence of the gene for dihydrolipoamide acetyltransferase
A;Reference number: A30198; MUID:89017234; PMID:3050999
A;Accession: A30198
A;Molecule type: mRNA
A;Residues: 1-482 <NIU>
A;Cross-references: UNIPROT:P12695; GB:J04096; NID:G170971; PIDN:AAA34385.1; PID:G170972
R;Poehlmann, R.; Philippsen, P.
submitted to the EMBL Data Library, April 1995
A;Reference number: S53896
A;Accession: S53909
A;Molecule type: DNA
A;Residues: 1-482 <POE>
A;Cross-references: EMBL:X86470; NID:G791101; PIDN:CAA60189.1; PID:G791115
R;Poehlmann, R.; Philippsen, P.
submitted to the Protein Sequence Database, April 1996
A;Reference number: S62997
A;Accession: S63003
A;Molecule type: DNA
A;Residues: 1-482 <POW>
A;Cross-references: EMBL:Z71347; NID:G1301954; PIDN:CAA95945.1; PID:G1301955; MIPS:YNL071
A;Experimental source: strain S288C
R;Poehlmann, R.; Philippsen, P.
Yeast 12, 391-402, 1996
A;Title: Sequencing a cosmid clone of *Saccharomyces cerevisiae* chromosome XIV reveals 12
A;Reference number: S63925; MUID:96267764; PMID:8701611
A;Accession: S63938
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-482 <POF>
A;Cross-references: EMBL:X86470; NID:G791101; PIDN:CAA60189.1; PID:G791115
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1995
C;Genetics:
A;Gene: SGD:LAT1; PDA2; ODP2
A;Cross-references: SGD:S0005015; MIPS:YNL071W
A;Map position: 14L
C;Superfamily: dihydrolipoamide acetyltransferase; lipoyl/biotin-binding homology
C;Keywords: acyltransferase; coenzyme A; mitochondrion
F;36-110/Domain: lipoyl/biotin-binding homology <LPB>

Query Match 61.2%; Score 41; DB 2; Length 482;
Best Local Similarity 58.3%; Pred. No. 66;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 SATTGSSPCP 12
|||:||||
Db 235 SSTAGSAPSPS 246

RESULT 11
T27155
hypothetical protein Y54E5A.7 - *Caenorhabditis elegans*
C;Species: *Caenorhabditis elegans*
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T27155
R;Lennard, N.
submitted to the EMBL Data Library, October 1998
A;Reference number: Z20320
A;Accession: T27155

A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-622 <WIL>
A;Cross-references: UNIPROT:Q9XWK4; EMBL:AL032643; PIDN:CAA21656.1; GSPDB:GN00019; CESP:
A;Experimental source: clone Y54ESA
C;Genetics:
A;Gene: CESP:Y54ESA.7
A;Map position: 1
A;Introns: 17/3; 39/3; 70/3; 225/3; 288/1; 349/1; 400/2; 480/3; 536/1; 579/3

Query Match 61.2%; Score 41; DB 2; Length 622;
Best Local Similarity 63.6%; Pred.No. 85;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 TTGTSSPGPTQ 13
:|||||:
Db 346 STTGRSPSTE 356

RESULT 12
T34369
hypothetical protein T19D12.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
C;Accession: T34369
R;Favella, A.
submitted to the EMBL Data Library, November 1995
A;Description: The sequence of C. elegans cosmid T19D12.
A;Reference number: Z21513
A;Accession: T34369
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1777 <FAV>
A;Cross-references: UNIPROT:Q22579; EMBL:U41263; PIDN:AAC24428.1; GSPDB:GN00020; CESP:T
A;Experimental source: strain Bristol N2; clone T19D12
C;Genetics:
A;Gene: CESP:T19D12.1
A;Map position: 2
A;Introns: 36/1; 134/2; 180/1; 622/3; 691/2; 754/1; 1111/2; 1174/1; 1271/3; 1322/2; 1681

Query Match 61.2%; Score 41; DB 2; Length 1777;
Best Local Similarity 70.0%; Pred.No. 2.Se+02;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 ATTTGSSPGP 11
:|||||:
Db 846 STTTGSTPAP 855

RESULT 13
S16162
cruzipain (EC 3.4.22.-) - Trypanosoma cruzi (fragment)
C;Species: Trypanosoma cruzi
C;Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 22-Jun-1999
C;Accession: S16162
R;Aelund, L.; Henriksson, J.; Campetella, O.; Frasch, A.C.C.; Pettersson, U.; Cazzulo, J
Mol. Biochem. Parasitol. 45, 345-348, 1991
A;Title: The C-terminal extension of the major cysteine proteinase (cruzipain) from Tryp
A;Reference number: S16162; MUID:91246270; PMID:2038364
A;Accession: S16162
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-173 <ASL>
A;Cross-references: EMBL:X54414; NID:g10605; PIDN:CAA38278.1; PID:g10606
C;Superfamily: papain
C;Keywords: cysteine proteinase; hydrolase

Query Match 59.7%; Score 40; DB 2; Length 173;
Best Local Similarity 58.3%; Pred.No. 33;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 SATTTGSSPGPT 12
:|||||:

Search completed: July 18, 2005, 13:41:52
Job time : 22.02 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2005, 12:59:21 ; Search time 98.54 Seconds
(without alignments)
67.557 Million cell updates/sec

Title: SEQ4

Perfect score: 67

Sequence: 1 sattuqsspgptq 13

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 03.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	63	94.0	513	1 GUX1_TRIKO	P62695 trichoderma
2	63	94.0	513	1 GUX1_TRIRE	P62694 trichoderma
3	63	94.0	514	2 G6UJY1	Q6UJY1 trichoderma
4	59	88.1	514	1 GUX1_TRIVI	P19355 trichoderma
5	55	82.1	302	1 AXE1_TRIRE	Q99034 trichoderma
6	47	70.1	526	2 Q7Q5T7	Q7Q5T7 anopheles 9
7	45	67.2	378	2 Q7TQK9	Q7TQK9 mus musculus
8	45	67.2	1552	2 Q68FW7	Q68FW7 mus musculus
9	44	65.7	435	2 Q89X12	Q89X12 bradyrhizob
10	44	65.7	663	1 RGP2_HUMAN	P47736 homo sapien
11	44	65.7	663	2 Q7Z5S8	Q7Z5S8 homo sapien
12	43	64.2	196	2 P95718	P95718 streptomyc
13	43	64.2	486	2 Q8T5C4	Q8T5C4 aedes aegypt
14	43	64.2	979	2 Q6ZP25	Q6ZP25 mus musculus
15	42	62.7	196	2 Q6J333	Q6J333 homo sapien
16	42	62.7	203	2 Q6J1T0	Q6J1T0 mus musculus
17	42	62.7	354	2 Q6PNA3	Q6PNA3 triticum ae
18	42	62.7	370	2 Q6ZV17	Q6ZV17 homo sapien
19	42	62.7	397	2 Q9ALN2	Q9ALN2 saccharopol
20	42	62.7	502	2 Q69WT1	Q69WT1 oryza sativ
21	42	62.7	519	1 RGS3_HUMAN	P49796 homo sapien
22	42	62.7	593	2 Q8MR60	Q8MR60 drosophila
23	42	62.7	664	2 Q7S749	Q7S749 neurospora
24	42	62.7	676	2 Q7TSD9	Q7TSD9 sudan ebola
25	42	62.7	693	2 Q6VTF3	Q6VTF3 choristoneu
26	42	62.7	733	2 Q8XQ09	Q8XQ09 neurospora
27	42	62.7	800	2 Q62237	Q62237 caenorhabdi
28	42	62.7	800	2 P90975	P90975 caenorhabdi
29	42	62.7	917	2 Q8IUQ1	Q8IUQ1 homo sapien
30	42	62.7	917	2 Q8NFN4	Q8NFN4 homo sapien
31	42	62.7	917	2 Q8WXA0	Q8WXA0 homo sapien

32 42 62.7 995 1 FOG1_MOUSE
33 42 62.7 1093 2 Q8NFN5
34 42 62.7 1198 2 Q6ZRM5
35 42 62.7 1209 2 Q7SZF4
36 42 62.7 1217 1 ITN1_RAT
37 42 62.7 1270 1 ITN1_XENLA
38 42 62.7 1424 2 Q9VL42
39 42 62.7 1714 1 ITN1_MOUSE
40 42 62.7 1721 1 ITN1_HUMAN
41 41 61.2 244 2 Q743G1
42 41 61.2 256 2 Q6QJ5
43 41 61.2 285 2 Q6KCY0
44 41 61.2 290 2 Q8CVV3
45 41 61.2 329 2 Q9BGY2

ALIGNMENTS

RESULT 1
GUX1 TRIKO STANDARD; PRT; 513 AA.
AC P62695: P00725;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Exoglucanase I precursor (EC 3.2.1.91) (Exocellobiohydrolase I) (CBHI)
GN Name=cbhl;
OS Trichoderma koningii.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
OX NCBI_TaxID=55202;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=G-39;
RX MEDLINE=94100788; PubMed=7764306;
RA Wey T.T., Hsue T.H., Huang L.;
RT "Molecular cloning and sequence analysis of the cellobiohydrolase I
Curr. Microbiol. 28:31-39(1994).
CC -I- FUNCTION: The biological conversion of cellulose to glucose
generally requires three types of hydrolytic enzymes: (1)
Endoglucanases which cut internal beta-1,4-glucosidic bonds; (2)
Exocellobiohydrolases that cut the disaccharide cellobiose from
the nonreducing end of the cellulose polymer chain; (3) Beta-1,4-
glucosidases which hydrolyze the cellobiose and other short cello-
oligosaccharides to glucose.
CC -I- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
in cellulose and cellotetraose, releasing cellobiose from the non-
reducing ends of the chains.
CC -I- SUBCELLULAR LOCATION: Secreted.
CC -I- SIMILARITY: Belongs to the glycosyl hydrolase 7 (cellulase C)
family.

CC -I- SIMILARITY: Contains 1 fungal-type cellulose-binding (CBD) domain.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

CC ENBL: X69976; CAA49596.1; -
DR InterPro: IPR000254; CBD_fungal.
DR InterPro: IPR008985; ConA_like lec_gl.
DR InterPro: IPR001722; Glyco_hydro_7.
DR Pfam: PF00734; CBM 1; 1.
DR Pfam: PF00840; Glyco_hydro_7; 1.
DR ProDom: PD001821; CBD_fungal; 1.
DR ProDom: PD186135; Glyco_hydro_7; 1.
DR SMART: SM00236; fCBG; 1.

```
DR PROSITE; PS00562; CBD_FUNGAL; 1.
KW Cellulose degradation; Glycoprotein; Glycosidase; Hydrolase;
KW Pyridone carboxylic acid; Signal.
FT SIGNAL 1 17 By similarity.
FT CHAIN 18 513 Exoglucanase I.
FT DOMAIN 18 453 Catalytic.
FT DOMAIN 454 477 Linker.
FT DOMAIN 478 513 Cellulose-binding.
FT ACT_SITE 143 143 By similarity.
FT ACT_SITE 229 229 Nucleophile (By similarity).
FT ACT_SITE 234 234 Proton donor (By similarity).
FT MOD_RES 18 18 Pyridone carboxylic acid (By
FT similarity).
FT CARBOHYD 62 62 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 81 81 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 287 287 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 401 401 N-linked (GlcNAc... ) (Potential).
FT DISULFID 21 89 By similarity.
FT DISULFID 36 42 By similarity.
FT DISULFID 67 88 By similarity.
FT DISULFID 78 84 By similarity.
FT DISULFID 155 414 By similarity.
FT DISULFID 189 227 By similarity.
FT DISULFID 193 226 By similarity.
FT DISULFID 247 273 By similarity.
FT DISULFID 255 260 By similarity.
FT DISULFID 278 348 By similarity.
FT DISULFID 485 502 By similarity.
FT DISULFID 496 512 By similarity.
SQ SEQUENCE 513 AA; 54073 MW; 9F5C0A8A54F2C12 CRC64;

Query Match 94.0%; Score 63; DB 1; Length 513;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ATTTGSSPGPTQ 13
Db 468 ATTTGSSPGPTQ 479
|||||
|||||

RESULT 2
GUX1 TRIRE STANDARD; PRT; 513 AA.
AC P62694; P00725;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Exoglucanase I precursor (EC 3.2.1.91) (Exocellobiohydrolase I) (CBHI)
DE (1,4-beta-cellobiohydrolase).
GN Name=cbhl;
OS Trichoderma reesei (Hypocrea jecorina).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
OX NCBI_TaxID=51453;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=L27;
RA Shoemaker S., Schweickart V., Ladner M., Gelfand D., Kwok S.,
RA Myambo K., Innis M.;
RT "Molecular cloning of exo-cellobiohydrolase I derived from Trichoderma
RT reesei strain L27.";
RL Biotechnology (N.Y.) 1:691-696(1983).
RN [2]
RP ACTIVE SITE.
RA Tomme P., Claysens M.;
RT "Identification of a functionally important carboxyl group in
RT cellobiohydrolase I from Trichoderma reesei.";
RL FEBS Lett. 243:239-243(1989).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.81 ANGSTROMS) OF 18-452.
RX MEDLINE=94310436; PubMed=8036499;
RA Divine C., Staahlberg J., Reinikainen T., Ruohonen L., Pettersson G.,
RA Knowles J.K.C., Teeri T.T., Jones T.A.;
```

```
RT "The three-dimensional crystal structure of the catalytic core of
RT cellobiohydrolase I from Trichoderma reesei.";
RL Science 265:524-528(1994).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 18-452.
RC STRAIN=QM9414 / Rut C-30;
RX MEDLINE=98128795; PubMed=9466911; DOI=10.1006/jmbi.1997.1437;
RA Divine C., Staahlberg J., Teeri T.T., Jones T.A.;
RT "High-resolution crystal structures reveal how a cellulose chain is
RT bound in the 50 A long tunnel of cellobiohydrolase I from Trichoderma
RT reesei.";
RL J. Mol. Biol. 275:309-325(1998).
RN [5]
RP STRUCTURE BY NMR OF 478-513.
RX MEDLINE=90057416; PubMed=2554967;
RA Kraulis P.J., Clore G.M., Nilges M., Jones T.A., Pettersson G.,
RA Knowles J., Gronenborn A.M.;
RT "Determination of the three-dimensional solution structure of the C-
RT terminal domain of cellobiohydrolase I from Trichoderma reesei. A
RT study using nuclear magnetic resonance and hybrid distance geometry-
RT dynamical simulated annealing.";
RL Biochemistry 28:7241-7257(1989).
RN [6]
RP STRUCTURE BY NMR OF 478-513.
RX MEDLINE=97194052; PubMed=9041630;
RA Mattinen M.L., Kontteli M., Kerovuo J., Linder M., Annala A.,
RA Lindeberg G., Reinikainen T., Drakenberg T.;
RT "Three-dimensional structures of three engineered cellulose-binding
RT domains of cellobiohydrolase I from Trichoderma reesei.";
RL Protein Sci. 6:294-303(1997).
CC -!- FUNCTION: The biological conversion of cellulose to glucose
CC generally requires three types of hydrolytic enzymes: (1)
CC Endoglucanases which cut internal beta-1,4-glucosidic bonds; (2)
CC Exocellobiohydrolases that cut the disaccharide cellobiose from
CC the nonreducing end of the cellulose polymer chain; (3) Beta-1,4-
CC glucosidases which hydrolyze the cellobiose and other short cello-
CC oligosaccharides to glucose.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC in cellulose and cellotetraose, releasing cellobiose from the non-
CC reducing ends of the chains.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: T.reesei produces two different
CC exocellobiohydrolases. They are unique in that they can hydrolyze
CC crystalline cellulose in the absence of endoglucanases.
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 7 (cellulase C)
CC family.
CC -!- SIMILARITY: Contains 1 fungal-type cellulose-binding (CBD) domain.
DR PIR; A09092; EUTQI.
DR PDB; 1AZ6; NMR; -.
DR PDB; 1AZH; NMR; -.
DR PDB; 1AZJ; NMR; -.
DR PDB; 1AZK; NMR; -.
DR PDB; 1CBH; NMR; -.
DR PDB; 1CEL; X-ray; -.
DR PDB; 1DY4; X-ray; -.
DR PDB; 1EGN; X-ray; -.
DR PDB; 2CBH; NMR; -.
DR PDB; 2CEL; X-ray; -.
DR PDB; 3CEL; X-ray; -.
DR PDB; 4CEL; X-ray; -.
DR PDB; 5CEL; X-ray; -.
DR PDB; 6CEL; X-ray; -.
DR PDB; 7CEL; X-ray; -.
DR PDB; 8CEL; Model; -.
DR InterPro; IPR000254; CBD_fungal.
DR InterPro; IPR008985; ConA_like lec gl.
DR InterPro; IPR001722; Glyco_hydro_7.
DR Pfam; PF00734; CBM 1; 1.
DR Pfam; PF00840; Glyco_hydro_7; 1.
DR ProDom; PD001821; CBD_fungal; 1.
DR ProDom; PD186135; Glyco_hydro_7; 1.
DR SMART; SM00236; fCBH; 1.
DR PROSITE; PS00562; CBD_FUNGAL; 1.
```

KW 3D-structure; Cellulose degradation; Glycoprotein; Glycosidase;
KW Hydrolase; Pyrrolidone carboxylic acid; Signal.

FT SIGNAL 1 17
FT CHAIN 18 513 Exoglucanase I.
FT DOMAIN 18 453 Catalytic.
FT DOMAIN 454 477 Linker.
FT DOMAIN 478 513 Cellulose-binding.
FT ACT_SITE 143 143 Probable.
FT ACT_SITE 229 229 Nucleophile.
FT ACT_SITE 234 234 Proton donor.
FT MOD_RES 18 18 Pyrrolidone carboxylic acid.
FT CARBOHYD 62 62 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 81 81 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 287 287 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 401 401 N-linked (GlcNAc...) (Potential).
FT DISULFID 21 89
FT DISULFID 36 42
FT DISULFID 67 88
FT DISULFID 78 84
FT DISULFID 155 414
FT DISULFID 189 227
FT DISULFID 193 226
FT DISULFID 247 273
FT DISULFID 255 260
FT DISULFID 278 348
FT DISULFID 485 502
FT DISULFID 496 512
FT STRAND 19 20
FT STRAND 30 36
FT TURN 38 39
FT STRAND 42 51
FT HELIX 53 55
FT STRAND 58 60
FT TURN 61 62
FT STRAND 66 66
FT STRAND 68 69
FT TURN 70 71
FT STRAND 72 73
FT TURN 75 77
FT HELIX 81 87
FT STRAND 88 90
FT TURN 95 99
FT STRAND 101 104
FT TURN 105 106
FT STRAND 107 111
FT STRAND 113 115
FT STRAND 119 121
FT STRAND 123 130
FT TURN 131 132
FT STRAND 133 134
FT STRAND 136 137
FT TURN 140 141
FT STRAND 142 149
FT TURN 151 152
FT TURN 155 156
FT STRAND 157 164
FT TURN 168 174
FT TURN 176 177
FT HELIX 182 184
FT TURN 185 185
FT STRAND 190 190
FT TURN 191 192
FT STRAND 199 200
FT TURN 201 202
FT STRAND 203 204
FT TURN 207 208
FT STRAND 210 211
FT TURN 216 217
FT STRAND 219 220
FT STRAND 223 226
FT STRAND 229 235
FT STRAND 240 245
FT STRAND 242 247

FT STRAND 253 256
FT HELIX 257 260
FT TURN 263 264
FT TURN 268 269
FT STRAND 273 273
FT STRAND 280 280
FT TURN 282 286
FT TURN 288 289
FT STRAND 291 292
FT TURN 295 296
FT STRAND 299 300
FT TURN 301 302
FT STRAND 305 311
FT TURN 313 314
FT STRAND 317 323
FT TURN 324 325
FT STRAND 326 329
FT STRAND 333 335
FT TURN 336 337
FT STRAND 338 340
FT STRAND 343 343
FT HELIX 345 355
FT HELIX 359 362

Query Match 94.0%; Score 63; DB 1; Length 513;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ATTTGSSPGPTQ 13
|||||
DB 468 ATTTGSSPGPTQ 479

RESULT 3

Q6UJY1 PRELIMINARY; PRT; 514 AA.
AC Q6UJY1; 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE Cellobiohydrolase I.
GN Name-cbhl;
OS Trichoderma viride.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; mitosporic Hypocreales; Trichoderma.
OX NCBI_TaxID=5547;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AS 3.3711;
RA Liu B.D., Yang Q., Zhou Q., Song J.Z.;
RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY368686; AAQ76092.1; -
DR HSSP; P56680; 1A39.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl...; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR000254; CBD fungal.
DR InterPro; IPR008985; ConA like lec_gl.
DR InterPro; IPR001722; Glyco_hydro_7.
DR Pfam; PF00734; CBM_1; 1.
DR Pfam; PF00840; Glyco_hydro_7; 1.
DR PRINTS; PR00734; GLHYDRLASE7.
DR ProDom; PD001821; CBD fungal; 1.
DR ProDom; PD186135; Glyco_hydro_7; 1.
DR SMART; SM00236; fCBD; 1.
DR PROSITE; PS00562; CBD_FUNGAL; 1.
KW Hydrolase.
SQ SEQUENCE 514 AA; 54041 MW; 5ED7D35962A07BA8 CRC64;

Query Match 94.0%; Score 63; DB 2; Length 514;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY      2 ATTGSSPGTQ 13
      |||||
Db      469 ATTGSSPGTQ 480

RESULT 4
GUX1_TRIVI STANDARD; PRT; 514 AA.
AC P19355; O93932;
DT 01-NOV-1990 (Rel. 16, Created)
DT 25-OCT-2004 (Rel. 45, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Exoglucanase I precursor (EC 3.2.1.91) (Exocellobiohydrolase) (1,4-
DE beta-cellobiohydrolase).
GN Names:cbhl;
OS Trichoderma viride.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; mitosporic Hypocreales; Trichoderma.
OX NCBI_TaxID=5547;
RN [1]_SEQUENCE FROM N.A.
RX MEDLINE=91016856; PubMed=2216737;
RA Cheng C.; Tsukagoshi N.; Uda K. S.;
RT "Nucleotide sequence of the cellobiohydrolase gene from Trichoderma
RT viride.";
RL Nucleic Acids Res. 18:5559-5559(1990).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=MC300-1;
RA Watanabe M.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: The biological conversion of cellulose to glucose
CC Generally requires three types of hydrolytic enzymes: (1)
CC Endoglucanases which cut internal beta-1,4-glucosidic bonds; (2)
CC Exocellobiohydrolases that cut the disaccharide cellobiose from
CC the nonreducing end of the cellulose polymer chain; (3) Beta-1,4-
CC glucosidases which hydrolyze the cellobiose and other short cello-
CC oligosaccharides to glucose.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC in cellulose and cellotetraose, releasing cellobiose from the non-
CC reducing ends of the chains.
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 7 (cellulase C)
CC family.
CC -!- SIMILARITY: Contains 1 fungal-type cellulose-binding (CBD) domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X53931; CAA37878.1; -
DR EMBL; AB021656; BAA36215.1; -
DR PIR; S11439; S11439.
DR HSSP; P00725; 6CEL.
DR InterPro; IPR000254; CBD_fungal.
DR InterPro; IPR008985; ConA like lec_gl.
DR InterPro; IPR001722; Glyco_hydro_7.
DR Pfam; PF00734; CBM_1; 1.
DR Pfam; PF00840; Glyco_hydro_7; 1.
DR ProDom; PD001821; CBD_fungal; 1.
DR ProDom; PD186135; Glyco_hydro_7; 1.
DR SMART; SM00236; fCBD; 1.
DR PROSITE; PS00562; CBD_FUNGAL; 1.
KW Cellulose degradation; Glycoprotein; Glycosidase; Hydrolase; Signal.
FT SIGNAL 1 17
FT CHAIN 18 514 Exoglucanase I.
FT DOMAIN 18 453 Catalytic.
FT DOMAIN 454 478 Linker.
FT DOMAIN 479 514 Cellulose-binding (By similarity).

```

```

FT ACT_SITE 229 Nucleophile (By similarity).
FT ACT_SITE 234 Proton donor (By similarity).
FT DISULFID 21 By similarity.
FT DISULFID 36 By similarity.
FT DISULFID 67 By similarity.
FT DISULFID 78 By similarity.
FT DISULFID 155 By similarity.
FT DISULFID 189 By similarity.
FT DISULFID 193 By similarity.
FT DISULFID 247 By similarity.
FT DISULFID 255 By similarity.
FT DISULFID 278 By similarity.
FT DISULFID 486 By similarity.
FT DISULFID 497 By similarity.
FT CARBOHYD 62 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 81 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 287 N-linked (GlcNAc...) (Potential).
FT CONFLICT 173 S -> T (in Ref. 1).
FT CONFLICT 256 D -> E (in Ref. 1).
FT CONFLICT 401 N -> D (in Ref. 1).
FT CONFLICT 414 C -> S (in Ref. 1).
FT CONFLICT 449 S -> P (in Ref. 1).
FT CONFLICT 462 Missing (in Ref. 1).
FT CONFLICT 466 R -> P (in Ref. 1).
FT CONFLICT 492 S -> I (in Ref. 1).
SQ SEQUENCE 514 AA; 54002 MW; C5B1FB734C9DAF5 CRC64;

Query Match 88.1%; Score 59; DB 1; Length 514;
Best Local Similarity 91.7%; Pred. No. 0.77;
Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 ATTGSSPGTQ 13
      |||||
Db      469 ATTGSSPGTQ 480

RESULT 5
AXE1_TRIRE STANDARD; PRT; 302 AA.
AC Q99034;
DT 29-MAR-2004 (Rel. 43, Created)
DT 29-MAR-2004 (Rel. 43, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Acetylxyylan esterase precursor (EC 3.1.1.72).
GN Name=axe1;
OS Trichoderma reesei (Hypocrea jecorina).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
OX NCBI_TaxID=51453;
RN [1]_SEQUENCE FROM N.A.; SEQUENCE OF 158-186, AND CHARACTERIZATION.
RP SEQUENCE FROM N.A.; Rut C-30;
RC STRAIN=QM9414 / Rut C-30;
RX MEDLINE=96235218; PubMed=8647098;
RA Margolles-Clark E.; Tenkanen M.; Soederlund H.; Penttilae M.;
RT "Acetyl xyylan esterase from Trichoderma reesei contains an active-site
RT serine residue and a cellulose-binding domain.";
RL Eur. J. Biochem. 237:553-560(1996).
RN [2]
RP FUNCTION.
RC STRAIN=QM9414 / Rut C-30;
RA Poutanen K.; Sundberg M.; Korte H.; Puls J.;
RT "Deacetylation of xylians by acetyl esterases of Trichoderma reesei.";
RL Appl. Microbiol. Biotechnol. 33:506-510(1990).
RN [3]
RP CHARACTERIZATION.
RC STRAIN=QM9414 / Rut C-30;
RA Sundberg M.; Poutanen K.;
RT "Purification and properties of two acetylxyylan esterases of
RT Trichoderma reesei";
RL Biotechnol. Appl. Biochem. 13:1-11(1991).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 32-238, AND MASS
RP SPECTROMETRY.

```

RX MEDLINE=98437545; PubMed=9761918; DOI=10.1107/S0907444997012213;
 RA Hakulinen N., Tenkanen M., Rouvinen J.;
 RT "Crystallization and preliminary X-ray diffraction studies of the
 RL catalytic core of acetyl xylan esterase from *Trichoderma reesei*."; Acta Crystallogr. D 54:430-432 (1998).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 32-238.
 RX PubMed=11243887; DOI=10.1006/j.sbi.2000.4318;
 RA Hakulinen N., Tenkanen M., Rouvinen J.;
 RT "Three-dimensional structure of the catalytic core of acetyl xylan
 RL esterase from *Trichoderma reesei*: insights into the deacetylation
 RT mechanism."; J. Struct. Biol. 132:180-190 (2000).
 RL J. Struct. Biol. 132:180-190 (2000).
 CC -!- FUNCTION: Degrades acetylated xyans by cleaving acetyl side
 CC groups from the hetero-xylan backbone.
 CC -!- CATALYTIC ACTIVITY: Deacetylation of xyans and xyloligosaccharides.
 CC -!- ENZYME REGULATION: Inhibited by phenylmethylsulfonyl fluoride.
 CC -!- PATHWAY: Xylan degradation.
 CC -!- SUBUNIT: Monomer.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- PTM: Glycosylated.
 CC -!- SIMILARITY: Contains 1 fungal-type cellulose-binding (CBD) domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; Z69256; CAA93247.1; -;
 DR PIR; S71334; S71334.
 DR PDB; 1QOZ; X-ray; A/B=-;
 DR InterPro; IPR000254; CBD_fungal.
 DR InterPro; IPR000675; Cutinase.
 DR InterPro; IPR008262; Lipase_AS.
 DR Pfam; PF00734; CBM 1; 1.
 DR Pfam; PF01083; Cutinase; 1.
 DR ProDom; PD001821; CBD_fungal; 1.
 DR SMART; SM00236; fCBD-1.
 DR PROSITE; PS00562; CBD_FUNGAL; FALSE NEG.
 DR PROSITE; PS00120; LIPASE_SER; UNKNOWN 1.
 DR 3D-structure; Cellulose degradation; Direct protein sequencing;
 KW Glycoprotein; Cellulase; Pyrolidone carboxylic acid; Serine esterase;
 KW Signal.
 FT SIGNAL 1 20 Potential.
 FT PROPEP 21 31 Potential.
 FT CHAIN 32 302 Acetyl xylan esterase.
 FT DOMAIN 244 266 Linker (By similarity).
 FT DOMAIN 267 302 Cellulose-binding (By similarity).
 FT MOD_RES 32 32 Pyridone carboxylic acid.
 FT ACT_SITE 121 121 By similarity.
 FT DISULFID 274 291 By similarity.
 FT DISULFID 285 301 By similarity.
 FT CARBOHYD 94 94 N-linked (GlcNAc...) (Probable).
 SQ SEQUENCE 302 AA; 30754 MW; BB6EDCA2971A9F2A CRC64;
 Query Match 82.1%; Score 55; DB 1; Length 302;
 Best Local Similarity 76.9%; Pred. No. 1.8;
 Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 SATTTGSSPGPTQ 13
 Db 256 TSRTGSSPGPTQ 268
 RESULT 6
 Q7Q517 PRELIMINARY; PRT; 526 AA.
 AC Q7Q517
 DT 01-MAR-2004 (TrEMBLrel. 26, Created)

DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE AGCP6090 (Fragment).
 GN Name=agCG52987; ORFNames=ENSANG00000017996;
 OS Anopheles gambiae str. PEST.
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
 OX NCBI_TaxID=180454;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PEST;
 RA Anopheles Genome Sequencing Consortium;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 CC -!- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 DR EMBL; AAB01008960; EAA10725.1; -;
 DR HSSP; Q13231; 1GUV.
 DR GO; GO:0005576; C:extracellular; IEA.
 DR GO; GO:0008061; F:chitin binding; IEA.
 DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR GO; GO:0006030; P:chitin metabolism; IEA.
 DR GO; GO:0008152; P:metabolism; IEA.
 DR InterPro; IPR011583; Chitinase_II.
 DR InterPro; IPR002557; Chitin bind_Pera.
 DR InterPro; IPR001223; Glyco_hydro_18.
 DR InterPro; IPR001579; Glyco_hydro_18AS.
 DR Pfam; PF01607; CBM 14; 1.
 DR Pfam; PF00704; Glyco_hydro_18; 1.
 DR ProDom; PD000471; Chitinase_II; 1.
 DR PROSITE; PS01095; CHITINASE_18; 1.
 DR PROSITE; PS0940; CHIT_BIND_II; 1.
 KW Glycosidase; Hydrolase.
 FT NON_TER 1
 SQ SEQUENCE 526 AA; 57099 MW; E87452250249245C CRC64;
 Query Match 70.1%; Score 47; DB 2; Length 526;
 Best Local Similarity 61.5%; Pred. No. 56;
 Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
 QY 1 SATTTGSSPGPTQ 13
 Db 451 TSRTGAPGPTQ 463
 RESULT 7
 Q7QK9 PRELIMINARY; PRT; 378 AA.
 AC Q7QK9
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Arhgef11 protein (Fragment).
 GN Name=Arhgef11;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Olfactory epithelium;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Klausner R.L., Feingold E.A., Grouse L.H., Dege J.G.,
 RA Strausberg R.L., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loughran N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

```

RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RC SEQUENCE FROM N.A.
RC TISSUE=Olfactory epithelium;
RA Strausberg R.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC054078; AAH54078.1; -.
FT NON_TER 1
SQ SEQUENCE 378 AA; 40560 MW; ACBDF3D1E2BFE98 CRC64;

Query Match 67.2%; Score 45; DB 2; Length 378;
Best Local Similarity 81.8%; Pred. No. 80;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 SATTTGSSPGP 11
Db ||||| |||||
368 SATDTASSPGP 378

RESULT 8
Q68FM7 PRELIMINARY; PRT; 1552 AA.
AC Q68FM7;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altshuler S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RC SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Director MGC Project;
RL Submitted (AUG-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC079565; AAH79565.1; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001849; PH.
DR InterPro; IPR011036; PH_related.

```

```

DR InterPro; IPR00342; RGS.
DR InterPro; IPR00219; RhogEF.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00621; RhogEF; 1.
DR SMART; SM00228; PDZ; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00315; RGS; 1.
DR SMART; SM00325; RhogEF; 1.
DR PROSITE; PS50010; DH_2; 1.
DR PROSITE; PS50106; PDZ; 1.
DR PROSITE; PS50003; PH DOMAIN; 1.
DR PROSITE; PS50132; RGS; 1.
KW Hypothetical protein.
SQ SEQUENCE 1552 AA; 171801 MW; 8FF6AFC41AD63258 CRC64;

Query Match 67.2%; Score 45; DB 2; Length 1552;
Best Local Similarity 81.8%; Pred. No. 3.6e+02;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 SATTTGSSPGP 11
Db ||||| |||||
1542 SATDTASSPGP 1552

RESULT 9
Q89XI2 PRELIMINARY; PRT; 435 AA.
AC Q89XI2;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE B110332 protein.
GN OrderedLocustNames=b110332;
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USD110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USD110.";
RL DNA Res. 9:189-197(2002).
DR EMBL; AP005936; BAC45597.1; -.
DR GO; GO:0030288; C:periplasmic space (sensu Gram-negative Bact. . .; IEA.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR009056; Cytochrome_c.
DR InterPro; IPR003088; Cyt C.
DR InterPro; IPR001311; SBP7glu receptor.
DR InterPro; IPR001638; SBP_bac_3.
DR Pfam; PF00034; Cytochrom_C; 1.
DR SMART; SM00062; PBPb; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 435 AA; 46789 MW; 7BC92E1744DD3C1E CRC64;

Query Match 65.7%; Score 44; DB 2; Length 435;
Best Local Similarity 61.5%; Pred. No. 1.3e+02;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 SATTTGSSPGPTQ 13
Db : ||||| : |||||
318 AGTTTGTGGVGTQ 330

```

```
RESULT 10
RGP2_HUMAN STANDARD; PRT; 663 AA.
ID P47736;
AC 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Rap1 GTPase-activating protein 1 (Rap1GAP).
GN Name=RAP1GAL; (Human).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=91256304; PubMed=1904317;
RA Rubinfield B., Munemitsu S., Clark R., Conroy L., Watt K.,
RA Crosier W.J., McCormick F., Polakis P.;
RT "Molecular cloning of a GTPase activating protein specific for the
RT Krev-1 protein p21rap1.";
RL Ccell 65:1033-1042(1991).
RN [2]
RP TISSUE SPECIFICITY. AND INDUCTION.
RX MEDLINE=98010656; PubMed=9345962; DOI=10.1074/jbc.272.44.28081;
RA Kurachi H., Wada Y., Tsukamoto N., Maeda M., Kubota H., Hattori M.,
RA Iwai K., Minato N.;
RT "Human SPA-1 product selectively expressed in lymphoid tissues is a
RT specific GTPase-activating protein for Rap1 and Rap2.";
RL J Biol. Chem. 272:28081-28088(1997).
CC -!- FUNCTION: GTPase activator for the nuclear Ras-related regulatory
CC protein RAP-1A (KREV-1), converting it to the putatively inactive
CC GDP-bound state.
CC -!- SUBCELLULAR LOCATION: Associated with Golgi membranes.
CC -!- TISSUE SPECIFICITY: Significant expression seen in the brain,
CC kidney and pancreas. Abundant in the cerebral cortex and expressed
CC at much lower levels in the spinal cord. Not detected in the
CC lymphoid tissues.
CC -!- INDUCTION: By 12-O-tetradecanoylphorbol-13-acetate (TPA) in
CC promyelocytic HL-60 cells.
CC -!- SIMILARITY: Contains 1 GoLoco domain.
CC -!- SIMILARITY: Contains 1 Rap-GAP domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M64788; AAA60252.1; -
DR PIR; A39897; A39897.
DR Genew; HGNC:9858; RAP1GAL.
DR MIM; 600278; -
DR GO; GO:0005096; F:GTPase activator activity; TAS.
DR GO; GO:0007165; P:signal transduction; TAS.
DR InterPro; IPR003109; GoLoco.
DR InterPro; IPR000331; Rap GAP.
DR Pfam; PF02188; GoLoco; 1.
DR Pfam; PF02145; Rap GAP; 1.
DR SMART; SM00390; GoLoco; 1.
DR PROSITE; PS50877; GOLOCO; 1.
DR PROSITE; PS50085; RAPGAP; 1.
DR PROSITE; PS50085; RAPGAP; 1.
KW GTPase activation; Membrane.
FT DOMAIN 1 17 GoLoco.
FT DOMAIN 181 397 Rap-GAP.
SQ SEQUENCE 663 AA; 73397 MW; 3703B7CC603404DA CRC64;

Query Match 65.7%; Score 44; DB 1; Length 663;
Best Local Similarity 61.5%; Pred. No. 2.1e+02;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

RESULT 12
P95718 PRELIMINARY; PRT; 196 AA.
ID P95718 PRELIMINARY; PRT; 196 AA.
AC P95718;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)

Query Match 65.7%; Score 44; DB 2; Length 663;
Best Local Similarity 61.5%; Pred. No. 2.1e+02;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Query Match 65.7%; Score 44; DB 2; Length 663;
Best Local Similarity 61.5%; Pred. No. 2.1e+02;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Query Match 65.7%; Score 44; DB 2; Length 663;
Best Local Similarity 61.5%; Pred. No. 2.1e+02;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
```

DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
 DE 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Ar2d protein.
 GN Name=ard2;
 OS Streptomyces capreolus.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Pseudonocardineae; Actinosynnemataceae; Saccharothrix.
 OX NCBI_TaxID=66854;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRRL3817;
 RX MEDLINE=95255257; PubMed=7737149;
 RA Barrasa M.I., Tercero J.A., Lacaille R.A., Jimenez A.;
 RT "The ard1 gene from Streptomyces capreolus encodes a polypeptide of
 the ABC-transporters superfamily which confers resistance to the
 aminonucleoside antibiotic A201A";
 RL Eur. J. Biochem. 228:562-569 (1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRRL3817;
 RX MEDLINE=97274639; PubMed=9128724;
 RA Barrasa M., Tercero J., Jimenez A.;
 RT "The aminonucleoside antibiotic A201A is inactivated by a
 phosphotransferase activity from Streptomyces capreolus NRRL 3817, the
 producing organism. Isolation and molecular characterization of the
 relevant encoding gene and its DNA flanking regions.";
 RL Eur. J. Biochem. 245:54-63 (1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRRL3817;
 RX MEDLINE=22310933; PubMed=12423351;
 RA Saugar I., Sanz E., Rubio M., Espinosa J., Jimenez A.;
 RT "Identification of a set of genes involved in the biosynthesis of the
 aminonucleoside moiety of antibiotic A201A from Streptomyces
 capreolus.";
 RL Eur. J. Biochem. 269:5527-5535 (2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRRL3817;
 RX MEDLINE=NRRL3817;
 RA Sanz E., Saugar I., Jimenez A.;
 RT Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRRL3817;
 RA Saugar I.;
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; X84374; CAD82197.1; -.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0000166; F:nucleotide binding; IEA.
 DR InterPro; IPR003593; AAA_ATPase.
 DR SMART; SM00382; AAA; 1.
 KW ATP-binding.
 SQ SEQUENCE 196 AA; 21736 MW; 28BD6D6132751554B CRC64;

 Query Match 64.2%; Score 43; DB 2; Length 196;
 Best Local Similarity 61.5%; Pred. No. 82;
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

 QY 1 SATTTGSSPGPTQ 13
 |||:|||||
 Db 181 SAVPTGSDPGPVR 193

 RESULT 13
 Q8T5C4 PRELIMINARY; PRT; 486 AA.
 AC Q8T5C4;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Peritrophin.
 OS Aedes aegypti (Yellowfever mosquito).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes.
 OX NCBI_TaxID=71159;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Shao L., Jacobs-Lorena M.;
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY050566; AAL05409.1; -.
 DR HSSP; P91818; 1DOC.
 DR GO; GO:0005576; C:extracellular; IEA.
 DR GO; GO:0008061; F:chitin binding; IEA.
 DR GO; GO:0006030; P:chitin metabolism; IEA.
 DR InterPro; IPR002557; Chitin_bind_PeRA.
 DR Pfam; PF01607; CBM_14; 5.
 DR SMART; SM00494; ChtBD2; 5.
 DR PROSITE; PS50940; CHTT_BIND II; 5.
 SQ SEQUENCE 486 AA; 52390 MW; 4D4BE418FE2FE404 CRC64;

 Query Match 64.2%; Score 43; DB 2; Length 486;
 Best Local Similarity 66.7%; Pred. No. 2.1e+02;
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

 QY 1 SATTTGSSPGPT 12
 |:::|||||
 Db 182 SSTTSSSPSPT 193

 RESULT 14
 Q8ZPZ5 PRELIMINARY; PRT; 979 AA.
 ID Q6ZPZ5
 AC Q6ZPZ5;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE MKIAA1055 protein (Fragment).
 GN Name=mkIAA1055;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryonic tail;
 RX PubMed=14621295;
 RA Okazaki N., Kikuno R., Inamoto S., Koseki H., Hiraoka S.,
 RA Suga Y., Nagase T., Ohara O., Koga H.;
 RT "Prediction of the coding sequences of mouse homologues of KIAA gene:
 RT cDNAs identified by screening of terminal sequences of 500 mouse KIAA-homologous
 RT randomly sampled from size-fractionated libraries.";
 RL DNA Res. 10:167-180 (2003).
 DR EMBL; AK129273; BAC98083.1; -.
 DR InterPro; IPR001849; PH.
 DR InterPro; IPR000195; RabGAP_TBC.
 DR Pfam; PF00566; TBC; 1.
 DR SMART; SM00233; PH; 1.
 DR SMART; SM00164; TBC; 1.
 DR PROSITE; PS50003; PH_DOMAIN; 1.
 DR PROSITE; PS50086; TBC_RABGAP; 1.
 FT NON_TER 1
 SQ SEQUENCE 979 AA; 111543 MW; AAC7843177E29357 CRC64;

 Query Match 64.2%; Score 43; DB 2; Length 979;
 Best Local Similarity 61.5%; Pred. No. 4.5e+02;
 Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

 QY 1 SATTTGSSPGPTQ 13
 :||:|||||
 Db 339 TATSGSSGPGTK 351

 RESULT 15
 Q6J333 PRELIMINARY; PRT; 196 AA.
 ID Q6J333

AC O6J333;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DE 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE Intersectin 1 isoform 11 (Fragment).
 GN Name=ITGN1;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Skrypkina I., Tsyba L., Rynditch A., Nikolaienko O., Gardiner K.;
 RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: Contains 1 SH3 domain.
 DR EMBL; AY603419; AAT28306.1; -.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00018; SH3 1; 1.
 DR ProDom; PD000066; SH3; 1.
 DR PROSITE; PS50002; SH3; 1.
 KW SH3 domain.
 FT NON_TER 1
 FT NON_TER 196
 SQ SEQUENCE 196 AA; 21151 MW; 24F852506CF76233 CRC64;

Query Match 62.7%; Score 42; DB 2; Length 196;
 Best Local Similarity 80.0%; Pred. No. 1.2e+02;
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ATTTGSSGPG 11
 || |||||
 Db 130 ATATGSSPSG 139

Search completed: July 18, 2005, 13:33:23
 Job time : 100.54 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2005, 12:59:21 ; Search time 109.46 Seconds
(without alignments)
45.934 Million cell updates/sec

Title: SEQ4

Perfect score: 67

Sequence: 1 satttgsspgptq 13

Scoring table: BLOSSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	67	100.0	13	3 AAY71089	Synthetic
2	63	94.0	24	2 AAR77260	Exo-cello
3	63	94.0	28	2 AAR21408	Sequence
4	63	94.0	28	2 AAR32568	CBH I hin
5	63	94.0	269	2 AAR32569	Fusion pr
6	63	94.0	485	8 ADK81798	Cellobioh
7	63	94.0	436	4 AAB352239	T reesei
8	63	94.0	436	4 AAB352237	T reesei
9	63	94.0	436	4 AAB352240	T reesei
10	63	94.0	496	8 ADK81865	Cellobioh
11	63	94.0	497	8 ADK81814	Cellobioh
12	63	94.0	437	8 ADK81820	Cellobioh
13	63	94.0	437	8 ADK81830	Cellobioh
14	63	94.0	437	8 ADK81861	Cellobioh
15	63	94.0	437	8 ADK81884	Cellobioh
16	63	94.0	437	8 ADK81784	Cellobioh
17	63	94.0	437	8 ADK81796	Cellobioh
18	63	94.0	437	8 ADK81816	Cellobioh
19	63	94.0	437	8 ADK81879	Cellobioh
20	63	94.0	437	8 ADK81812	Cellobioh
21	63	94.0	437	8 ADK81878	Cellobioh
22	63	94.0	437	8 ADK81800	Cellobioh
23	63	94.0	437	8 ADK81828	Cellobioh
24	63	94.0	437	8 ADK81863	Cellobioh
25	63	94.0	437	8 ADK81863	Cellobioh

26	63	94.0	497	8 ADK81868	Cellobioh
27	63	94.0	497	8 ADK81893	Cellobioh
28	63	94.0	497	8 ADK81817	Cellobioh
29	63	94.0	497	8 ADK81825	Cellobioh
30	63	94.0	497	8 ADK81846	Cellobioh
31	63	94.0	497	8 ADK81847	Cellobioh
32	63	94.0	497	8 ADK81849	Cellobioh
33	63	94.0	497	8 ADK81854	Cellobioh
34	63	94.0	497	8 ADK81786	Cellobioh
35	63	94.0	497	8 ADK81813	Cellobioh
36	63	94.0	497	8 ADK81831	Cellobioh
37	63	94.0	497	8 ADK81834	Cellobioh
38	63	94.0	497	8 ADK81867	Cellobioh
39	63	94.0	497	8 ADK81875	Cellobioh
40	63	94.0	497	8 ADK81780	Cellobioh
41	63	94.0	497	8 ADK81808	Cellobioh
42	63	94.0	497	8 ADK81819	Cellobioh
43	63	94.0	497	8 ADK81838	Cellobioh
44	63	94.0	497	8 ADK81844	Cellobioh
45	63	94.0	497	8 ADK81850	Cellobioh

ALIGNMENTS

RESULT 1

AAY71089
ID AAY71089 standard; peptide; 13 AA.

XX AAY71089;

XX 21-SEP-2000 (first entry)

DE Synthetic linker peptide #4 encoded by MV07JA oligonucleotide linker.

XX Ilama: HC-V; heavy chain variable domain; antigen binding protein;
linker; conformational flexibility; multivalent binding protein; bi-head;
human chorionic gonadotropin; hCG; immunoassay; agglutination assay;
purification.

OS Synthetic.

XX Key Location/Qualifiers

PH Peptide 2..10

FT /label= Peptide linker 4

FT /note= "Flanked by one residue from N- and C-terminii of

FT HCV fragment"

XX WO200024884-A2.

XX 04-MAY-2000.

XX 22-OCT-1999; 99WO-EP008323.

XX 27-OCT-1998; 98WO-EP006991.

XX 22-APR-1999; 99EP-00303118.

XX (UNIL) UNILEVER PLC.

XX (UNIL) UNILEVER NV.

XX (HIND-) HINDUSTAN LEVER LTD.

XX Frenken LGJ, Howell S, Van Der Vaart JM;

XX WPI; 2000-350728/30.

XX N-PSDB; AAD00663.

XX Use of a linker whose amino acid sequence confers restricted

XX conformational flexibility to generate multivalent and multispecific

XX antigen binding proteins.

XX Example 1.1d; Page 20; 50pp; English.

XX The present sequence is the synthetic linker peptide #4, encoded by the

CC oligonucleotide linker fragment, MW07JA. It consists of the last residue
 CC of the N-terminal HC-V fragment (S) and the first residue of the C-
 CC terminal HC-V fragment (O), intersected by the connecting linker peptide.
 CC It is used for the construction of *Saccharomyces cerevisiae* episomal
 CC expression plasmid, pUKS333, encoding anti-hCG-anti-RR6 bispecific
 CC biheads, containing the linker peptide. The peptide linker confers
 CC restricted conformational flexibility for linking binding units in a
 CC multivalent binding protein. The linker is used to generate multivalent
 CC or multispecific antigen binding proteins for immunoassays, agglutination
 CC assays or for purification
 XX
 SQ Sequence 13 AA;

Query Match 100.0%; Score 67; DB 3; Length 13;
 Best Local Similarity 100.0%; Pred. No. 0.0024;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SATTTGSSPGPTQ 13
 |||||
 Db 1 SATTTGSSPGPTQ 13

RESULT 2
 AAR77260
 ID AAR77260 standard; protein; 24 AA.

XX AAR77260;
 XX
 XX 25-MAR-2003 (revised)
 DT 13-DEC-1995 (first entry)
 XX
 XX Exo-cellobiohydrolase CBHI hinge region.
 DE
 XX Cellulase; hinge region; enzyme.

XX Trichoderma longibrachiatum.

XX WO9516782-A1.

XX 22-JUN-1995.

XX 19-DEC-1994; 94WO-US014163.

XX 17-DEC-1993; 93US-00169948.

XX (GEMV) GENENCOR INT INC.

XX Fowler T, Clarkson KA, Ward M, Collier KD, Larenas E;

XX WPI; 1995-231574/30.

XX N-PSDB; AAQ91280.

XX Pure, truncated fungal cellulase protein from *Trichoderma* - useful to
 PT reduce or eliminate dye, colourant or pigment back-staining or
 PT redeposition in stone-washing or bio-polishing.

XX Claim 22; Page 54; 105pp; English.

XX A DNA gene fragment (AAQ91279) derived from *Trichoderma* which encodes for
 CC the CBHI catalytic core protein is claimed. The encoded protein
 CC (AAR77259) is capable of exhibiting exoglucanase activity. The DNA
 CC fragment may be operably linked to a DNA fragment coding for a hinge
 CC region DNA. A DNA fragment comprising AAQ91279 and AAQ91280 operably
 CC linked is claimed. Genes for CBHI and CBHI have been isolated from *T.*
 CC longibrachiatum and the protein domain structure has been confirmed
 CC (Shoemaker, S. et al. 1983, *Bio/Technology*, 1, 691-696; Teeri, T. et al.
 CC 1983, *Bio/Technology* 1, 696-699 and Teeri, T. et al., 1987, *Gene*, 51, 43-
 CC 52). (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 24 AA;

Query Match 94.0%; Score 63; DB 2; Length 24;
 Best Local Similarity 100.0%; Pred. No. 0.018;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 ATTTGSSPGPTQ 13
 |||||
 Db 10 ATTTGSSPGPTQ 21

RESULT 3

AAR21408
 ID AAR21408 standard; peptide; 28 AA.

XX AAR21408;

XX
 XX
 DT 09-JAN-2003 (revised)

DT 17-MAY-1992 (first entry)

XX Sequence of a single chain antibody, having cbh1 hinge as a linker
 DE encoded by pML5.

XX Immunoglobulin; fusion protein; heavy chain; light chain.

XX Synthetic.

XX WO9201797-A.

XX 06-FEB-1992.

XX 16-JUL-1990; 90US-00552757.

XX 16-JUL-1990; 90US-00552757.

XX (ALKO-) ALKO OY.

XX Nyyssonen E, Keranen S, Penttila M, Takkinen K, Knowles JKC;

XX WPI; 1992-064953/08.

XX N-PSDB; AAQ21395.

XX Immunoglobulin gene-contg. vector for transformation of *trichoderma* -
 PT recombinant proteins useful in diagnosis and control of physiological and
 PT agricultural conditions.

XX Example; Fig 29; 138pp; English.

XX The inventors claim a fusion protein (FP) comprising immunologically
 CC active fragments of light and heavy chains linked to each other by
 CC synthetic, *Trichoderma* derived or heterologous linkers. The FP comprises
 CC a variable domain of the light chain or heavy chain. The FPs are useful
 CC in diagnosis and control of physiological and agricultural conditions.
 CC (Updated on 09-JAN-2003 to add missing OS field.)

XX Sequence 28 AA;

Query Match 94.0%; Score 63; DB 2; Length 28;

Best Local Similarity 100.0%; Pred. No. 0.022;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ATTTGSSPGPTQ 13

Db 14 ATTTGSSPGPTQ 25

RESULT 4

AAR32568
 ID AAR32568 standard; protein; 28 AA.

XX AAR32568;

XX 25-MAR-2003 (revised)

DT 08-JUN-1993 (first entry)

XX CBH I hinge linker peptide.

XX

PT New variant cellobiohydrolase (CBH) I cellulase from *Hypocrea jecorina*,
xx

PT useful as a detergent or feed additive for treating wood pulp or
 PT converting biomass to sugar.
 XX
 PS Disclosure; SEQ ID NO 30; 104pp; English.
 XX
 CC The invention describes a variant cellobiohydrolase (CBH) I cellulase
 CC from Hypocrea jecorina. A variant cellobiohydrolase (CBH) I cellulase
 CC from Hypocrea jecorina comprises a substitution or deletion at a position
 CC corresponding to one or more of residues S8, Q17, G22, T41, N49, S57,
 CC N64, A68, A77, N89, S92, N103, A112, S113, E193, S196, M213, L225, T226,
 CC P227, T246, D249, R251, Y252, T255, D257, S278, S279, K286, L288,
 CC E295, T296, S297, A299, N301, E325, T332, F338, S342, T356, Y371,
 CC T380, Y381, V393, R394, S398, V403, S411, G430, G440, T445, T462, T484,
 CC Q487, and P491 in CBH I. Also described are: a nucleic acid encoding the
 CC CBH I variant; a vector comprising the nucleic acid; a host cell
 CC transformed with the vector; a method of producing the CBH I variant by
 CC culturing the host cell in a suitable culture medium under conditions to
 CC produce CBH I variant, and obtaining the CBH I variant produced; a
 CC detergent composition comprising the CBH I variant; a method of treating wood
 CC pulp by contacting the wood pulp with a CBH I variant; and a method of
 CC converting biomass to sugars by contacting the biomass with a CBH I
 CC variant. The variant CBH I is useful as a laundry or dish detergent or
 CC feed additive for treating wood pulp or converting biomass to sugar. The
 CC enzyme is also disclosed as being potentially useful in the production of
 CC biofuels. This is the amino acid sequence of Hypocrea jecorina
 CC cellobiohydrolase 1 (CBH1 or exoglucanase 1) mutant T380G/Y381D/R394A
 CC delta381-393.
 XX
 SQ Sequence 485 AA;

Query Match 94.0%; Score 63; DB 8; Length 485;
 Best Local Similarity 100.0%; Pred. No. 0.43;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ATTTGSSPGPTQ 13
 |||||
 Db 440 ATTTGSSPGPTQ 451

RESULT 7
 AAB35239
 ID AAB35239 standard; protein; 496 AA.
 AC AAB35239;
 XX
 DT 24-APR-2001 (first entry)
 XX
 DE T reesei cellobiohydrolase I reduced glycosylation variant CBH1N384A.
 XX
 KW Cellobiohydrolase I reduced glycosylation variant; exoglucanase;
 KW CBH1N45A; CBH1N270A; CBH1N384A; CBH1; mutant; mutein.
 XX
 OS Hypocrea jecorina.
 OS Synthetic.
 XX
 PN WO200104284-A1.
 XX
 PD 18-JAN-2001.

PF 13-JUL-2000; 2000WO-US019007.
 XX
 PR 13-JUL-1999; 99US-0143711P.
 XX
 XX (MIDE) MIDWEST RES INST.
 PA
 PI Adney WS, Decker SR, Lantz McCarter S, Baker JO, Nieves R;
 PI Himmel ME, Vinzant TB;
 XX
 DR WPI; 2001-138339/14.

XX Making active exoglucanase in eukaryotic heterologous host, involves
 PT reducing glycosylation of exoglucanase by replacing N-glycosylation site

PT amino acid residue with non-glycosyl accepting amino acid residue.
 XX
 PS Claim 7; Page 16-18; 21pp; English.
 XX
 CC The present invention describes a method of producing an active
 CC exoglucanase in a eukaryotic host, involving reducing glycosylation of
 CC the enzyme, and provides the sequences of 3 modified Trichoderma reesei
 CC cellobiohydrolase 1 (CBH1) proteins, designated CBH1N45A, CBH1N270A and
 CC CBH1N384A. The exoglucanases are useful for pretreating biomass for the
 CC the efficient production of ethanol. The present sequence is the variant
 CC protein CBH1N384A
 XX
 SQ Sequence 496 AA;

Query Match 94.0%; Score 63; DB 4; Length 496;
 Best Local Similarity 100.0%; Pred. No. 0.44;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ATTTGSSPGPTQ 13
 |||||
 Db 451 ATTTGSSPGPTQ 462

RESULT 8
 AAB35238
 ID AAB35238 standard; protein; 496 AA.
 XX
 AC AAB35238;

XX
 DT 24-APR-2001 (first entry)
 XX
 DE T reesei cellobiohydrolase I reduced glycosylation variant CBH1N270A.
 XX
 KW Cellobiohydrolase I reduced glycosylation variant; exoglucanase;
 KW CBH1N45A; CBH1N270A; CBH1N384A; CBH1; mutant; mutein.

OS Hypocrea jecorina.
 OS Synthetic.
 XX
 PN WO200104284-A1.
 XX
 PD 18-JAN-2001.

PF 13-JUL-2000; 2000WO-US019007.

PR 13-JUL-1999; 99US-0143711P.

XX (MIDE) MIDWEST RES INST.

PI Adney WS, Decker SR, Lantz McCarter S, Baker JO, Nieves R;

PI Himmel ME, Vinzant TB;

DR WPI; 2001-138339/14.

XX Making active exoglucanase in eukaryotic heterologous host, involves
 PT reducing glycosylation of exoglucanase by replacing N-glycosylation site
 PT amino acid residue with non-glycosyl accepting amino acid residue.

PS Claim 6; Page 14-16; 21pp; English.

XX The present invention describes a method of producing an active
 CC exoglucanase in a eukaryotic host, involving reducing glycosylation of
 CC the enzyme, and provides the sequences of 3 modified Trichoderma reesei
 CC cellobiohydrolase 1 (CBH1) proteins, designated CBH1N45A, CBH1N270A and
 CC CBH1N384A. The exoglucanases are useful for pretreating biomass for the
 CC the efficient production of ethanol. The present sequence is the variant
 CC protein CBH1N270A
 XX

SQ Sequence 496 AA;

Query Match 94.0%; Score 63; DB 4; Length 496;
 Best Local Similarity 100.0%; Pred. No. 0.44;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY      2 ATTTGSSPGPTQ 13
Db      451 ATTTGSSPGPTQ 462

RESULT 9
AAB35237
ID AAB35237 standard; protein; 496 AA.
XX
AC AAB35237;
XX
DT 24-APR-2001 (first entry)
XX
DE T reesei cellobiohydrolase I reduced glycosylation variant CBH1N45A.
XX
KW Cellobiohydrolase I reduced glycosylation variant; exoglucanase;
KW CBH1N45A; CBH1N270A; CBH1N384A; CBH1; mutant; mutein.
XX
OS Hypocrea jecorina.
OS Synthetic.
XX
PN WO200104284-A1.
XX
PD 18-JAN-2001.
XX
PF 13-JUL-2000; 2000WO-US019007.
XX
PR 13-JUL-1999; 99US-0143711P.
XX
PA (WIDE ) MIDWEST RES INST.
XX
XX Adney WS, Decker SR, Lantz Mccarter S, Baker JO, Nieves R;
PI Himmel ME, Vinzant TB;
XX
DR WPI; 2001-138339/14.
XX
XX Making active exoglucanase in eukaryotic heterologous host, involves
PT reducing glycosylation of exoglucanase by replacing N-glycosylation site
PT amino acid residue with non-glycosyl accepting amino acid residue.
XX
PS Claim 5; Page 12-14; 21pp; English.
XX
CC The present invention describes a method of producing an active
CC exoglucanase in a eukaryotic host, involving reducing glycosylation of
CC the enzyme, and provides the sequences of 3 modified Trichoderma reesei
CC cellobiohydrolase 1 (CBH1) proteins, designated CBH1N45A, CBH1N270A and
CC CBH1N384A. The exoglucanases are useful for pretreating biomass for the
CC the efficient production of ethanol. The present sequence is the variant
CC protein CBH1N45A
XX
SQ Sequence 496 AA;

Query Match          94.0%; Score 63; DB 4; Length 496;
Best Local Similarity 100.0%; Pred. No. 0.44;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 ATTTGSSPGPTQ 13
Db      451 ATTTGSSPGPTQ 462

RESULT 10
AAB35240
ID AAB35240 standard; protein; 496 AA.
XX
AC AAB35240;
XX
XX 11-SEP-2003 (revised)
DT 24-APR-2001 (first entry)
XX
DE T reesei cellobiohydrolase I CBH1.
XX

```

```

KW Cellobiohydrolase I reduced glycosylation variant; exoglucanase;
KW CBH1N45A; CBH1N270A; CBH1N384A; CBH1; mutant; mutein.
XX
OS Hypocrea jecorina.
XX
PN WO200104284-A1.
XX
PD 18-JAN-2001.
XX
PF 13-JUL-2000; 2000WO-US019007.
XX
PR 13-JUL-1999; 99US-0143711P.
XX
PA (WIDE ) MIDWEST RES INST.
XX
PI Adney WS, Decker SR, Lantz Mccarter S, Baker JO, Nieves R;
PI Himmel ME, Vinzant TB;
XX
DR WPI; 2001-138339/14.
XX
XX Making active exoglucanase in eukaryotic heterologous host, involves
PT reducing glycosylation of exoglucanase by replacing N-glycosylation site
PT amino acid residue with non-glycosyl accepting amino acid residue.
XX
PS Example 1; Page 18-20; 21pp; English.
XX
CC The present invention describes a method of producing an active
CC exoglucanase in a eukaryotic host, involving reducing glycosylation of
CC the enzyme, and provides the sequences of 3 modified Trichoderma reesei
CC cellobiohydrolase 1 (CBH1) proteins, designated CBH1N45A, CBH1N270A and
CC CBH1N384A. The exoglucanases are useful for pretreating biomass for the
CC the efficient production of ethanol. The present sequence is the CBH1
CC protein. (Updated on 11-SEP-2003 to standardise OS field)
XX
SQ Sequence 496 AA;

Query Match          94.0%; Score 63; DB 4; Length 496;
Best Local Similarity 100.0%; Pred. No. 0.44;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 ATTTGSSPGPTQ 13
Db      451 ATTTGSSPGPTQ 462

RESULT 11
ADK81865
ID ADK81865 standard; protein; 496 AA.
XX
AC ADK81865;
XX
XX 20-MAY-2004 (first entry)
DT
XX
DE Cellobiohydrolase I (CBH1) mutant T41I delta T445.
XX
KW cellobiohydrolase I; CBH I; cellulase; detergent; feed additive;
KW wood pulp treatment; biomass conversion; sugar production; biofuel;
KW exoglucanase 1; mutant; mutein.
XX
XX Hypocrea jecorina.
OS
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 41 /note= "wild type Thr substituted by Ile"
XX
PN WO2004016760-A2.
XX
PD 26-FEB-2004.
XX
PF 15-AUG-2003; 2003WO-US025625.
XX
PR 16-AUG-2002; 2002US-0404063P.
XX

```

```

PR 21-MAR-2003; 2003US-0456368P.
PR 27-MAR-2003; 2003US-0458696P.
PR 27-MAR-2003; 2003US-0458853P.
XX
XX (GEMV ) GENENCOR INT INC.
XX
XX Day AG, Goedegebuur F, Gualfetti P, Mitchinson C, Neefe P;
XX Sandgren M, Shaw A, Stahlberg J;
XX
XX WPI; 2004-257195/24.
XX
XX New variant cellobiohydrolase (CBH) I cellulase from Hypocrea jecorina,
XX useful as a detergent or feed additive for treating wood pulp or
XX converting biomass to sugar.
XX
XX Claim 5; Page; 104pp; English.
XX
XX The invention describes a variant cellobiohydrolase (CBH) I cellulase
XX from Hypocrea jecorina. A variant cellobiohydrolase (CBH) I cellulase
XX from Hypocrea jecorina comprises a substitution or deletion at a position
XX corresponding to one or more of residues S8, Q17, G22, T41, N49, S57,
XX N64, A68, A77, N89, S92, N103, A112, S113, E193, S196, M213, L225, T226,
XX P227, T246, D249, R251, Y252, T255, D257, S278, S279, K286, L288,
XX E295, T296, S297, R394, S398, V403, S411, G430, T445, T462, T484,
XX T380, Y381, V393, R394, S398, V403, S411, G430, T445, T462, T484,
XX T380, Y381, V393, R394, S398, V403, S411, G430, T445, T462, T484,
XX Q487, and P491 in CBH I. Also described are: a nucleic acid encoding the
XX CBH I variant; a vector comprising the nucleic acid; a host cell
XX transformed with the vector; a method of producing the CBH I variant by
XX culturing the host cell in a suitable culture medium under conditions to
XX produce CBH I variant, and obtaining the CBH I variant produced; a
XX detergent composition comprising a surfactant and the CBH I variant; a
XX feed additive comprising the CBH I variant; a method of treating wood
XX pulp by contacting the wood pulp with a CBH I variant; and a method of
XX converting biomass to sugars by contacting the biomass with a CBH I
XX variant. The variant CBH I is useful as a laundry or dish detergent or
XX feed additive for treating wood pulp or converting biomass to sugar. The
XX enzyme is also disclosed as being potentially useful in the production of
XX biofuels. This is the amino acid sequence of a Hypocrea jecorina
XX cellobiohydrolase 1 (CBH1 or exoglucanase 1) mutant. Note: This sequence
XX does not appear in the printed specification but has been created by the
XX indexer using information given in the invention.
XX
XX Sequence 496 AA;
XX
XX Query Match 94.0%; Score 63; DB 8; Length 496;
XX Best Local Similarity 100.0%; Pred. NO. 0.44;
XX Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 2 ATTTGSSPGPTQ 13
XX |||||
XX Db 451 ATTTGSSPGPTQ 462
XX
XX RESULT 12
XX ADK81814
XX ID ADK81814 standard; protein; 497 AA.
XX
XX AC ADK81814;
XX
XX XX 20-MAY-2004 (first entry)
XX
XX DE Cellobiohydrolase I (CBH1) mutant S92T.
XX
XX KW cellobiohydrolase I; CBH I; cellulase; detergent; feed additive;
XX wood pulp treatment; biomass conversion; sugar production; biofuel;
XX exoglucanase 1; mutant; mutein.
XX
XX OS Hypocrea jecorina.
XX OS Synthetic.
XX
XX PH Location/Qualifiers
XX FT Misc-difference 92 /note= "Wild type Ser substituted by Thr"
XX

```

```

XX
XX PN WO2004016760-A2.
XX
XX PD 26-FEB-2004.
XX
XX XX 15-AUG-2003; 2003WO-US025625.
XX
XX PF 16-AUG-2002; 2002US-0404063P.
XX
XX PR 21-MAR-2003; 2003US-0456368P.
XX
XX PR 27-MAR-2003; 2003US-0458696P.
XX
XX PR 27-MAR-2003; 2003US-0458853P.
XX
XX (GEMV ) GENENCOR INT INC.
XX
XX PA Day AG, Goedegebuur F, Gualfetti P, Mitchinson C, Neefe P;
XX Sandgren M, Shaw A, Stahlberg J;
XX
XX WPI; 2004-257195/24.
XX
XX DR New variant cellobiohydrolase (CBH) I cellulase from Hypocrea jecorina,
XX useful as a detergent or feed additive for treating wood pulp or
XX converting biomass to sugar.
XX
XX Claim 2; Page; 104pp; English.
XX
XX The invention describes a variant cellobiohydrolase (CBH) I cellulase
XX from Hypocrea jecorina. A variant cellobiohydrolase (CBH) I cellulase
XX from Hypocrea jecorina comprises a substitution or deletion at a position
XX corresponding to one or more of residues S8, Q17, G22, T41, N49, S57,
XX N64, A68, A77, N89, S92, N103, A112, S113, E193, S196, M213, L225, T226,
XX P227, T246, D249, R251, Y252, T255, D257, S278, S279, K286, L288,
XX E295, T296, S297, R394, S398, V403, S411, G430, T445, T462, T484,
XX T380, Y381, V393, R394, S398, V403, S411, G430, T445, T462, T484,
XX Q487, and P491 in CBH I. Also described are: a nucleic acid encoding the
XX CBH I variant; a vector comprising the nucleic acid; a host cell
XX transformed with the vector; a method of producing the CBH I variant by
XX culturing the host cell in a suitable culture medium under conditions to
XX produce CBH I variant, and obtaining the CBH I variant produced; a
XX detergent composition comprising a surfactant and the CBH I variant; a
XX feed additive comprising the CBH I variant; a method of treating wood
XX pulp by contacting the wood pulp with a CBH I variant; and a method of
XX converting biomass to sugars by contacting the biomass with a CBH I
XX variant. The variant CBH I is useful as a laundry or dish detergent or
XX feed additive for treating wood pulp or converting biomass to sugar. The
XX enzyme is also disclosed as being potentially useful in the production of
XX biofuels. This is the amino acid sequence of a Hypocrea jecorina
XX cellobiohydrolase 1 (CBH1 or exoglucanase 1) mutant. Note: This sequence
XX does not appear in the printed specification but has been created by the
XX indexer using information given in the invention.
XX
XX Sequence 497 AA;
XX
XX Query Match 94.0%; Score 63; DB 8; Length 497;
XX Best Local Similarity 100.0%; Pred. NO. 0.44;
XX Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 2 ATTTGSSPGPTQ 13
XX |||||
XX Db 452 ATTTGSSPGPTQ 463
XX
XX RESULT 13
XX ADK81820
XX ID ADK81820 standard; protein; 497 AA.
XX
XX AC ADK81820;
XX
XX XX 20-MAY-2004 (first entry)
XX
XX DE Cellobiohydrolase I (CBH1) mutant M213I.
XX
XX KW cellobiohydrolase I; CBH I; cellulase; detergent; feed additive;
XX wood pulp treatment; biomass conversion; sugar production; biofuel;
XX

```


exoglucanase 1; mutant; mutein.

Hypocrea jecorina.

Synthetic.

Key Location/Qualifiers

Misc-difference 213 /note= "Wild type Met substituted by Ile"

WO2004016760-A2.

26-FEB-2004.

15-AUG-2003; 2003WO-US025625.

16-AUG-2002; 2002US-0404063P.

21-MAR-2003; 2003US-0456368P.

27-MAR-2003; 2003US-0458696P.

27-MAR-2003; 2003US-0458853P.

(GEMV) GENENCOR INT INC.

Day AG, Goedegebuur F, Gualfetti P, Mitchinson C, Neefe P; Sandgren M, Shaw A, Stahlberg J;

WPI; 2004-257195/24.

New variant cellobiohydrolase (CBH) I cellulase from Hypocrea jecorina, useful as a detergent or feed additive for treating wood pulp or converting biomass to sugar.

Claim 2; Page; 104pp; English.

The invention describes a variant cellobiohydrolase (CBH) I cellulase from Hypocrea jecorina. A variant cellobiohydrolase (CBH) I cellulase from Hypocrea jecorina comprises a substitution or deletion at a position corresponding to one or more of residues S8, Q17, G22, T41, N49, S57, N64, A68, A77, N89, S92, NIQ3, A112, S113, E193, S196, M213, L225, T226, P227, T246, D249, R251, Y252, T255, D257, D259, S278, K286, L288, E295, T296, S297, A299, N301, E325, T332, F338, S342, F352, T356, Y371, T380, Y381, V393, R394, S398, V403, S411, G430, G440, T445, T462, T484, Q487, and P491 in CBH I. Also described are: a nucleic acid encoding the CBH I variant; a vector comprising the nucleic acid; a host cell transformed with the vector; a method of producing the CBH I variant by culturing the host cell in a suitable culture medium under conditions to produce CBH I variant, and obtaining the CBH I variant produced; a detergent composition comprising a surfactant and the CBH I variant; a feed additive comprising the CBH I variant; a method of treating wood pulp by contacting the wood pulp with a CBH I variant; and a method of converting biomass to sugars by contacting the biomass with a CBH I variant. The variant CBH I is useful as a laundry or dish detergent or feed additive for treating wood pulp or converting biomass to sugar. The enzyme is also disclosed as being potentially useful in the production of biofuels. This is the amino acid sequence of a Hypocrea jecorina cellobiohydrolase 1 (CBH I or exoglucanase 1) mutant. Note: This sequence does not appear in the printed specification but has been created by the indexer using information given in the invention.

Sequence 497 AA;

Query Match 94.0%; Score 63; DB 8; Length 497;

Best Local Similarity 100.0%; Pred. No. 0.44;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 ATTTGSSPGPTQ 13

452 ATTTGSSPGPTQ 463

RESULT 14

ADK81830

ID ADK81830 standard; protein; 497 AA.

XX

Db 452 ATTTGSSPGPTQ 463
|||||
RESULT 15
ADK81861
ID ADK81861 standard; protein; 497 AA.
XX
AC ADK81861;
XX
DT 20-MAY-2004 (first entry)
XX
DE Cellobiohydrolase I (CBH1) mutant A112E/T226A.
XX
KW cellobiohydrolase I; CBH I; cellulase; detergent; feed additive;
KW wood pulp treatment; biomass conversion; sugar production; biofuel;
KW exoglucanase I; mutant; mutein.
XX
OS Hypocrea jecorina.
OS Synthetic.
XX
PH Key Location/Qualifiers
FT Misc-difference 112
FT /note= "Wild type Ala substituted by Glu"
FT Misc-difference 226
FT /note= "Wild type Thr substituted by Ala"
XX
PN WO2004016760-A2.
XX
PD 26-FEB-2004.
XX
PF 15-AUG-2003; 2003WO-US025625.
XX
PR 16-AUG-2002; 2002US-0404063P.
PR 21-MAR-2003; 2003US-0456368P.
PR 27-MAR-2003; 2003US-0458696P.
PR 27-MAR-2003; 2003US-0458853P.
XX
PA (GEMV) GENENCOR INT INC.
XX
PI Day AG, Goedegebuur F, Gualfetti P, Mitchinson C, Neefe P;
PI Sandgren M, Shaw A, Stahlberg J;
XX
DR WPI; 2004-257195/24.
XX
PT New variant cellobiohydrolase (CBH) I cellulase from Hypocrea jecorina,
PT useful as a detergent or feed additive for treating wood pulp or
PT converting biomass to sugar.
XX
PS Claim 5; Page; 104pp; English.
XX
CC The invention describes a variant cellobiohydrolase (CBH) I cellulase
CC from Hypocrea jecorina. A variant cellobiohydrolase (CBH) I cellulase
CC from Hypocrea jecorina comprises a substitution or deletion at a position
CC corresponding to one or more of residues S8, Q17, G22, T41, N49, S57,
CC N64, A68, A77, N89, S92, N103, A112, S113, E193, S196, M213, L225, T226,
CC P227, T246, D249, R251, Y252, T255, D257, D259, S278, S279, K286, L288,
CC E295, T296, S297, A299, N301, E325, T332, F338, S342, F352, T356, Y371,
CC T380, Y381, V393, R394, S398, V403, S411, G430, G440, T445, T462, T484,
CC Q487, and P491 in CBH I. Also described are: a nucleic acid encoding the
CC CBH I variant; a vector comprising the nucleic acid; a host cell
CC transformed with the vector; a method of producing the CBH I variant by
CC culturing the host cell in a suitable culture medium under conditions to
CC produce CBH I variant, and obtaining the CBH I variant produced; a
CC detergent composition comprising a surfactant and the CBH I variant; a
CC feed additive comprising the CBH I variant; a method of treating wood
CC pulp by contacting the wood pulp with a CBH I variant; and a method of
CC converting biomass to sugars by contacting the biomass with a CBH I
CC variant. The variant CBH I is useful as a laundry or dish detergent or
CC feed additive for treating wood pulp or converting biomass to sugar. The
CC enzyme is also disclosed as being potentially useful in the production of
CC biofuels. This is the amino acid sequence of a Hypocrea jecorina
CC cellobiohydrolase 1 (CBH1 or exoglucanase 1) mutant. Note: This sequence

CC does not appear in the printed specification but has been created by the
CC indexer using information given in the invention.
XX
SQ Sequence 497 AA;
Query Match 94.0%; Score 63; DB 8; Length 497;
Best Local Similarity 100.0%; Pred. No. 0.44;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 ATTTGSSPGPTQ 13
|||||
DB 452 ATTTGSSPGPTQ 463
Search completed: July 18, 2005, 13:40:29
Job time : 109.46 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2005, 13:09:46 ; Search time 15.4 Seconds
(without alignments)
62.478 Million cell updates/sec

Title: SEQ5
Perfect score: 51
Sequence: 1 sanhsngasq 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79.*
1: Piri:**
2: Piri:**
3: Piri:**
4: Piri:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	70.6	298	2 B70309	hypothetical prote
2	35	68.6	341	2 S69802	N-acetylmuramoyl-L
3	35	68.6	418	2 G70864	probable transmem
4	35	68.6	495	1 MNXRSA	nonstructural prot
5	35	68.6	634	1 S35574	transcription fact
6	35	68.6	862	2 A96778	hypothetical prote
7	35	68.6	1234	2 T31623	hypothetical prote
8	34	66.7	179	2 S74894	hypothetical prote
9	34	66.7	241	2 S29697	C-8 sterol isomera
10	34	66.7	414	2 S36838	acetyl-CoA C-acylt
11	34	66.7	1107	2 AC0976	probable autotrans
12	34	66.7	1336	2 S41794	SBC3 protein - yea
13	34	66.7	1589	1 RGBY05	cell division cont
14	33	64.7	286	2 S72384	hypothetical prote
15	33	64.7	326	2 B47172	ficolin-beta - pig
16	33	64.7	345	2 T32273	hypothetical prote
17	33	64.7	369	2 C88030	protein F46F5.10 l
18	33	64.7	411	2 T22095	hypothetical prote
19	33	64.7	460	2 T43224	hypothetical prote
20	33	64.7	556	1 JDLV64	DNA-directed DNA p
21	33	64.7	732	2 T50143	pumilio family pro
22	33	64.7	791	2 B46171	coat protein - lei
23	33	64.7	884	1 JDLV59	DNA-directed DNA p
24	33	64.7	5175	2 T20992	hypothetical prote
25	33	64.7	5198	2 T43230	hemikentlin precurs
26	32	62.7	136	2 T31545	hypothetical prote
27	32	62.7	148	2 T26309	hypothetical prote
28	32	62.7	224	2 A81655	ribosomal protein
29	32	62.7	224	2 H71506	probable S3 riboso

30	32	62.7	265	2 T18668	hypothetical prote
31	32	62.7	302	2 T21843	hypothetical prote
32	32	62.7	333	2 I84743	hepatoma transmem
33	32	62.7	336	2 I49766	hepatoma transmem
34	32	62.7	364	2 G86340	protein F2D10.35 l
35	32	62.7	384	2 S76163	hypothetical prote
36	32	62.7	438	2 G69971	citrate transporter
37	32	62.7	470	2 F91033	hypothetical prote
38	32	62.7	470	2 G85877	hypothetical prote
39	32	62.7	517	2 AH2174	hypothetical prote
40	32	62.7	557	2 C84146	ABC transporter re
41	32	62.7	718	1 S33168	gene pointed prote
42	32	62.7	728	2 S45403	hypothetical prote
43	32	62.7	824	2 T21675	hypothetical prote
44	32	62.7	1200	2 S77524	chromosome segrega
45	32	62.7	1242	2 T39453	probable mrna atab

ALIGNMENTS

RESULT 1

B70309
hypothetical protein aq_091 - Aquifex aeolicus
C;Species: Aquifex aeolicus

C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 09-Jul-2004
C;Accession: B70309
C;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov

V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666; PMID:9537320
A;Accession: B70309

A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-298 <AQF>

A;Cross-references: UNIPROT:O66501; GB:AE000673; NID:g2982834; PIDN:AAC06468.1; PID:g298
A;Experimental source: strain VF5
C;Genetics:

C;Superfamily: Aquifex aeolicus hypothetical protein aq_091
Query Match 70.6%; Score 36; DB 2; Length 298;
Best Local Similarity 75.0%; Pred. No. 16;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 NHSGNASQ 10
|||
Db 146 NHNGNASK 153

RESULT 2

S69802
N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28), 36.5K - phage A511
N;Alternate names: endolysin

C;Species: phage A511
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 09-Jul-2004
C;Accession: S69802
R;Loessner, M.J.; Wendlinger, G.; Scherer, S.
Mol. Microbiol. 16, 1231-1241, 1995

A;Title: Heterogeneous endolysins in *Listeria monocytogenes* bacteriophages: a new class
A;Reference number: S69798; MUID:96020653; PMID:8577256
A;Accession: S69802

A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-341 <LOE>

A;Cross-references: UNIPROT:Q38653; EMBL:X85010; NID:g853748; PIDN:CAAS9368.1; PID:g8537
C;Genetics:
A;Gene: ply511
C;Keywords: hydrolase

Query Match 68.6%; Score 35; DB 2; Length 341;
Best Local Similarity 77.8%; Pred. No. 29;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 ANHSGNASQ 10
||:|||||

Db 231 ANYSGTASQ 239

RESULT 3
G70864
probable transmembrane transport protein - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Jul-2004
C;Accession: G70864
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: G70864
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-418 <COL>
A;Cross-references: UNIPROT:O53183; GB:AL021246; GB:AL123456; NID:g3261507; PIDN:CAA1603
A;Experimental source: strain H37RV
C;Genetics:
A;Gene: RV2456C

Query Match 68.6%; Score 35; DB 2; Length 418;
Best Local Similarity 85.7%; Pred. No. 37;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 NHSGNAS 9
||:|||||

Db 146 NHAGNAS 152

RESULT 4
MNXRSA
nonstructural protein NCPV2 - simian rotavirus Sall
N;Alternate names: nonstructural protein NS53
C;Species: simian rotavirus Sall
C;Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 09-Jul-2004
C;Accession: S08215
R;Mitchell, D.B.; Both, G.W.
Virology 174, 618-621, 1990
A;Title: Conservation of a potential metal binding motif despite extensive sequence divergence
A;Reference number: S08215; MUID:90163231; PMID:2154894
A;Accession: S08215
A;Molecule type: genomic RNA
A;Residues: 1-495 <MIT>
A;Cross-references: UNIPROT:P15687; EMBL:X14914; NID:g61889; PIDN:CAA33039.1; PID:g61890
C;Genetics:
A;Map position: segment 5
C;Superfamily: bovine rotavirus nonstructural protein NCPV2
C;Keywords: nonstructural protein; zinc finger

Query Match 68.6%; Score 35; DB 1; Length 495;
Best Local Similarity 66.7%; Pred. No. 44;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 SANHSGNAS 9
:|||||

Db 311 ASNHGPNAS 319

RESULT 5
S35574
transcription factor HNF-1A - chicken
N;Alternate names: hepatic nuclear factor 1-alpha; transcription factor LF-B1
C;Species: Gallus gallus (chicken)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004

C;Accession: S35574; S33229
R;Hoerlein, A.; Grajer, K.H.; Igo-Kemenes, T.
Biol. Chem. Hoppe-Seyler 374, 419-425, 1993
A;Title: Genomic structure of the POU-related hepatic transcription factor HNF-1-alpha.
A;Reference number: S35574; MUID:94030658; PMID:8105803
A;Accession: S35574
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-634 <HOE>
A;Cross-references: UNIPROT:Q90867; EMBL:X67690
R;Grajer, K.H.; Hoerlein, A.; Igo-Kemenes, T.
Biol. Chem. Hoppe-Seyler 374, 319-326, 1993
A;Title: Hepatic nuclear factor 1alpha (HNF-1alpha) is expressed in the oviduct of hens and in the oviduct of ducks.
A;Reference number: S33229; MUID:93332647; PMID:7687846
A;Accession: S33229
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 375-411 <GRA>
C;Genetics:
A;Introns: 179/1; 241/2; 321/1; 373/3; 441/1; 505/1; 545/3; 593/1
C;Complex: homodimer; can also form heterodimers with, for example, HNF-1B
C;Function:
A;Description: transcription activator required for the expression of a number of liver-specific genes
A;Note: also expressed in some other tissues, where it may play other roles
C;Superfamily: transcription factor HNF-1; homeobox homology
C;Keywords: DNA binding; homeobox; liver; nucleus; transcription regulation
F;1-33/Region: dimerization
F;225-280/Domain: homeobox homology <HOX>

Query Match 68.6%; Score 35; DB 1; Length 634;
Best Local Similarity 66.7%; Pred. No. 57;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 SANHSGNAS 9
||:|||||

Db 336 SSNHRGNSS 344

RESULT 6
A96778
hypothetical protein F9E10.30 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C;Accession: A96778
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, J.; Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Dewar, K.; Dhar, S.; Eickbush, S.H.; Eickbush, D.; Fierman, M.; Gaitanaris, G.; Gill, G.; Glazebrook, J.; Hake, C.; Harber, P.; Hieshima, G.; Holt, D.; Hunter, J.; Inou, K.; Iwano, W.; Jaiswal, R.; Jones, R.; Karp, P.; Kavanagh, T.; Kohn, J.; Kunkel, L.; Landry, M.; Llave, J.; Luo, Z.; Ma, L.; Mayhew, M.; Miller, L.; Mitchell, J.; Nelson, D.; Nishizawa, K.; O'Connor, K.; O'Neill, M.; O'Neill, S.; O'Neill, T.; O'Neill, W.; O'Neill, X.; O'Neill, Y.; O'Neill, Z.; O'Neill, A.; O'Neill, B.; O'Neill, C.; O'Neill, D.; O'Neill, E.; O'Neill, F.; O'Neill, G.; O'Neill, H.; O'Neill, I.; O'Neill, J.; O'Neill, K.; O'Neill, L.; O'Neill, M.; O'Neill, N.; O'Neill, O.; O'Neill, P.; O'Neill, Q.; O'Neill, R.; O'Neill, S.; O'Neill, T.; O'Neill, U.; O'Neill, V.; O'Neill, W.; O'Neill, X.; O'Neill, Y.; O'Neill, Z.; O'Neill, A.; O'Neill, B.; O'Neill, C.; O'Neill, D.; O'Neill, E.; O'Neill, F.; O'Neill, G.; O'Neill, H.; O'Neill, I.; O'Neill, J.; O'Neill, K.; O'Neill, L.; O'Neill, M.; O'Neill, N.; O'Neill, O.; O'Neill, P.; O'Neill, Q.; O'Neill, R.; O'Neill, S.; O'Neill, T.; O'Neill, U.; O'Neill, V.; O'Neill, W.; O'Neill, X.; O'Neill, Y.; O'Neill, Z.; O'Neill, A.; O'Neill, B.; O'Neill, C.; O'Neill, D.; O'Neill, E.; O'Neill, F.; O'Neill, G.; O'Neill, H.; O'Neill, I.; O'Neill, J.; O'Neill, K.; O'Neill, L.; O'Neill, M.; O'Neill, N.; O'Neill, O.; O'Neill, P.; O'Neill, Q.; O'Neill, R.; O'Neill, S.; O'Neill, T.; O'Neill, U.; O'Neill, V.; O'Neill, W.; O'Neill, X.; O'Neill, Y.; O'Neill, Z.; O'Neill, A.; O'Neill, B.; O'Neill, C.; O'Neill, D.; O'Neill, E.; O'Neill, F.; O'Neill, G.; O'Neill, H.; O'Neill, I.; O'Neill, J.; O'Neill, K.; O'Neill, L.; O'Neill, M.; O'Neill, N.; O'Neill, O.; O'Neill, P.; O'Neill, Q.; O'Neill, R.; O'Neill, S.; O'Neill, T.; O'Neill, U.; O'Neill, V.; O'Neill, W.; O'Neill, X.; O'Neill, Y.; O'Neill, Z.; O'Neill, A.; O'Neill, B.; O'Neill, C.; O'Neill, D.; O'Neill, E.; O'Neill, F.; O'Neill, G.; O'Neill, H.; O'Neill, I.; O'Neill, J.; O'Neill, K.; O'Neill, L.; O'Neill, M.; O'Neill, N.; O'Neill, O.; O'Neill, P.; O'Neill, Q.; O'Neill, R.; O'Neill, S.; O'Neill, T.; O'Neill, U.; O'Neill, V.; O'Neill, W.; O'Neill, X.; O'Neill, Y.; O'Neill, Z.; O'Neill, A.; O'Neill, B.; O'Neill, C.; O'Neill, D.; O'Neill, E.; O'Neill, F.; O'Neill, G.; O'Neill, H.; O'Neill, I.; O'Neill, J.; O'Neill, K.; O'Neill, L.; O'Neill, M.; O'Neill, N.; O'Neill, O.; O'Neill, P.; O'Neill, Q.; O'Neill, R.; O'Neill, S.; O'Neill, T.; O'Neill, U.; O'Neill, V.; O'Neill, W.; O'Neill, X.; O'Neill, Y.; O'Neill, Z.; O'Neill, A.; O'Neill, B.; O'Neill, C.; O'Neill, D.; O'Neill, E.; O'Neill, F.; O'Neill, G.; O'Neill, H.; O'Neill, I.; O'Neill, J.; O'Neill, K.; O'Neill, L.; O'Neill, M.; O'Neill, N.; O'Neill, O.; O'Neill, P.; O'Neill, Q.; O'Neill, R.; O'Neill, S.; O'Neill, T.; O'Neill, U.; O'Neill, V.; O'Neill, W.; O'Neill, X.; O'Neill, Y.; O'Neill, Z.; O'Neill, A.; O'Neill, B.; O'Neill, C.; O'Neill, D.; O'Neill, E.; O'Neill, F.; O'Neill, G.; O'Neill, H.; O'Neill, I.; O'Neill, J.; O'Neill, K.; O'Neill, L.; O'Neill, M.; O'Neill, N.; O'Neill, O.; O'Neill, P.; O'Neill, Q.; O'Neill, R.; O'Neill, S.; O'Neill, T.; O'Neill, U.; O'Neill, V.; O'Neill, W.; O'Neill, X.; O'Neill, Y.; O'Neill, Z.; O'Neill, A.; O'Neill, B.; O'Neill, C.; O'Neill, D.; O'Neill, E.; O'Neill, F.; O'Neill, G.; O'Neill, H.; O'Neill, I.; O'Neill, J.; O'Neill, K.; O'Neill, L.; O'Neill, M.; O'Neill, N.; O'Neill, O.; O'Neill, P.; O'Neill, Q.; O'Neill, R.; O'Neill, S.; O'Neill, T.; O'Neill, U.; O'Neill, V.; O'Neill, W.; O'Neill, X.; O'Neill, Y.; O'Neill, Z.; O'Neill, A.; O'Neill, B.; O'Neill, C.; O'Neill, D.; O'Neill, E.; O'Neill, F.; O'Neill, G.; O'Neill, H.; O'Neill, I.; O'Neill, J.; O'Neill, K.; O'Neill, L.; O'Neill, M.; O'Neill, N.; O'Neill, O.; O'Neill, P.; O'Neill, Q.; O'Neill, R.; O'Neill, S.; O'Neill, T.; O'Neill, U.; O'Neill, V.; O'Neill, W.; O'Neill, X.; O'Neill, Y.; O'Neill, Z.; O'Neill, A.; O'Neill, B.; O'Neill, C.; O'Neill, D.; O'Neill, E.; O'Neill, F.; O'Neill, G.; O'Neill, H.; O'Neill, I.; O'Neill, J.; O'Neill, K.; O'Neill, L.; O'Neill, M.; O'Neill, N.; O'Neill, O.; O'Neill, P.; O'Neill, Q.; O'Neill, R.; O'Neill, S.; O'Neill, T.; O'Neill, U.; O'Neill, V.; O'Neill, W.; O'Neill, X.; O'Neill, Y.; O'Neill, Z.; O'Neill, A.; O'Neill, B.; O'Neill, C.; O'Neill, D.; O'Neill, E.; O'Neill, F.; O'Neill, G.; O'Neill, H.; O'Neill, I.; O'Neill, J.; O'Neill, K.; O'Neill, L.; O'Neill, M.; O'Neill, N.; O'Neill, O.; O'Neill, P.; O'Neill, Q.; O'Neill, R.; O'Neill, S.; O'Neill, T.; O'Neill, U.; O'Neill, V.; O'Neill, W.; O'Neill, X.; O'Neill, Y.; O'Neill, Z.; O'Neill, A.; O'Neill, B.; O'Neill, C.; O'Neill, D.; O'Neill, E.; O'Neill, F.; O'Neill, G.; O'Neill, H.; O'Neill, I.; O'Neill, J.; O'Neill, K.; O'Neill, L.; O'Neill, M.; O'Neill, N.; O'Neill, O.; O'Neill, P.; O'Neill, Q.; O'Neill, R.; O'Neill, S.; O'Neill, T.; O'Neill, U.; O'Neill, V.; O'Neill, W.; O'Neill, X.; O'Neill, Y.; O'Neill, Z.; O'Neill, A.; O'Neill, B.; O'Neill, C.; O'Neill, D.; O'Neill, E.; O'Neill, F.; O'Neill, G.; O'Neill, H.; O'Neill, I.; O'Neill, J.; O'Neill, K.; O'Neill, L.; O'Neill, M.; O'Neill, N.; O'Neill, O.; O'Neill, P.; O'Neill, Q.; O'Neill, R.; O'Neill, S.; O'Neill, T.; O'Neill, U.; O'Neill, V.; O'Neill, W.; O'Neill, X.; O'Neill, Y.; O'Neill, Z.; O'Neill, A.; O'Neill, B.; O'Neill, C.; O'Neill, D.; O'Neill, E.; O'Neill, F.; O'Neill, G.; O'Neill, H.; O'Neill, I.; O'Neill, J.; O'Neill, K.; O'Neill, L.; O'Neill, M.; O'Neill, N.; O'Neill, O.; O'Neill, P.; O'Neill, Q.; O'Neill, R.; O'Neill, S.; O'Neill, T.; O'Neill, U.; O'Neill, V.; O'Neill, W.; O'Neill, X.; O'Neill, Y.; O'Neill, Z.; O'Neill, A.; O'Neill, B.; O'Neill, C.; O'Neill, D.; O'Neill, E.; O'Neill, F.; O'Neill, G.; O'Neill, H.; O'Neill, I.; O'Neill, J.; O'Neill, K.; O'Neill, L.; O'Neill, M.; O'Neill, N.; O'Neill, O.; O'Neill, P.; O'Neill, Q.; O'Neill, R.; O'Neill, S.; O'Neill, T.; O'Neill, U.; O'Neill, V.; O'Neill, W.; O'Neill, X.; O'Neill, Y.; O'Neill, Z.; O'Neill, A.; O'Neill, B.; O'Neill, C.; O'Neill, D.; O'Neill, E.; O'Neill, F.; O'Neill, G.; O'Neill, H.; O'Neill, I.; O'Neill, J.; O'Neill, K.; O'Neill, L.; O'Neill, M.; O'Neill, N.; O'Neill, O.; O'Neill, P.; O'Neill, Q.; O'Neill, R.; O'Neill, S.; O'Neill, T.; O'Neill, U.; O'Neill, V.; O'Neill, W.; O'Neill, X.; O'Neill, Y.; O'Neill, Z.; O'Neill, A.; O'Neill, B.; O'Neill, C.; O'Neill, D.; O'Neill, E.; O'Neill, F.; O'Neill, G.; O'Neill, H.; O'Neill, I.; O'Neill, J.; O'Neill, K.; O'Neill, L.; O'Neill, M.; O'Neill, N.; O'Neill, O.; O'Neill, P.; O'Neill, Q.; O'Neill, R.; O'Neill, S.; O'Neill, T.; O'Neill, U.; O'Neill, V.; O'Neill, W.; O'Neill, X.; O'Neill, Y.; O'Neill, Z.; O'Neill, A.; O'Neill, B.; O'Neill, C.; O'Neill, D.; O'Neill, E.; O'Neill, F.; O'Neill, G.; O'Neill, H.; O'Neill, I.; O'Neill, J.; O'Neill, K.; O'Neill, L.; O'Neill, M.; O'Neill, N.; O'Neill, O.; O'Neill, P.; O'Neill, Q.; O'Neill, R.; O'Neill, S.; O'Neill, T.; O'Neill, U.; O'Neill, V.; O'Neill, W.; O'Neill, X.; O'Neill, Y.; O'Neill, Z.; O'Neill, A.; O'Neill, B.; O'Neill, C.; O'Neill, D.; O'Neill, E.; O'Neill, F.; O'Neill, G.; O'Neill, H.; O'Neill, I.; O'Neill, J.; O'Neill, K.; O'Neill, L.; O'Neill, M.; O'Neill, N.; O'Neill, O.; O'Neill, P.; O'Neill, Q.; O'Neill, R.; O'Neill, S.; O'Neill, T.; O'Neill, U.; O'Neill, V.; O'Neill, W.; O'Neill, X.; O'Neill, Y.; O'Neill, Z.; O'Neill, A.; O'Neill, B.; O'Neill, C.; O'Neill, D.; O'Neill, E.; O'Neill, F.; O'Neill, G.; O'Neill, H.; O'Neill, I.; O'Neill, J.; O'Neill, K.; O'Neill, L.; O'Neill, M.; O'Neill, N.; O'Neill, O.; O'Neill, P.; O'Neill, Q.; O'Neill, R.; O'Neill, S.; O'Neill, T.; O'Neill, U.; O'Neill, V.; O'Neill, W.; O'Neill, X.; O'Neill, Y.; O'Neill, Z.; O'Neill, A.; O'Neill, B.; O'Neill, C.; O'Neill, D.; O'Neill, E.; O'Neill, F.; O'Neill, G.; O'Neill, H.; O'Neill, I.; O'Neill, J.; O'Neill, K.; O'Neill, L.; O'Neill, M.; O'Neill, N.; O'Neill, O.; O'Neill, P.; O'Neill, Q.; O'Neill, R.; O'Neill, S.; O'Neill, T.; O'Neill, U.; O'Neill, V.; O'Neill, W.; O'Neill, X.; O'Neill, Y.; O'Neill, Z.; O'Neill, A.; O'Neill, B.; O'Neill, C.; O'Neill, D.; O'Neill, E.; O'Neill, F.; O'Neill, G.; O'Neill, H.; O'Neill, I.; O'Neill, J.; O'Neill, K.; O'Neill, L.; O'Neill, M.; O'Neill, N.; O'Neill, O.; O'Neill, P.; O'Neill, Q.; O'Neill, R.; O'Neill, S.; O'Neill, T.; O'Neill, U.; O'Neill, V.; O'Neill, W.; O'Neill, X.; O'Neill, Y.; O'Neill, Z;

```

RESULT 7
T31623
hypothetical protein Y57A10A.a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T31623
R:Smyle, R.
submitted to the EMBL Data Library, September 1999
A:Reference number: Z21048
A:Accession: T31623
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1234 <WIL>
A:Cross-references: EMBL:AL117195; NID:61549729; PIDN:CA855006.1; CESP:Y57A10A.a
A:Experimental source: Clone Y57A10A
C:Genetics:
A:Introns: 32/2; 66/2; 555/2; 662/2; 701/2; 735/2; 774/2; 813/2; 852/2; 891/2; 928/2; 1000/2
Query Match 68.6%; Score 35; DB 2; Length 1234;
Best Local Similarity 66.7%; Pred. No. 1.2e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 SANHSGNAS 9
Db 943 TSNHSGNES 951

RESULT 8
S74894
hypothetical protein slr1261 - Synechocystis sp. (strain PCC 6803)
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
C:Accession: S74894
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
s.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S74894
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-179 <KAN>
A:Cross-references: UNIPROT:P73801; EMBL:D90909; GB:AB001339; NID:g1652844; PIDN:BAAL1785
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Superfamily: hypothetical protein slr1261
Query Match 66.7%; Score 34; DB 2; Length 179;
Best Local Similarity 70.0%; Pred. No. 23;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 SANHSGNASQ 10
Db 71 SSNRSGNALQ 80

RESULT 9
S29697
C-8 sterol isomerase - smut fungus (Ustilago maydis)
C:Species: Ustilago maydis (corn smut)
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 09-Jul-2004
C:Accession: S29697
R:Burden, R.S.; James, C.S.; Bailey, A.M.; Keon, J.P.; Croxon, R.; Bard, M.; Hargreaves,
submitted to the EMBL Data Library, October 1992
A:Description: Isolation and characterization of the ERG2 gene, encoding delta 8 - delta
A:Reference number: S29697
A:Accession: S29697
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-241 <BUR>
A:Cross-references: UNIPROT:P32360; EMBL:Z17311; NID:g29666; PID:g2967

```

```

C:Genetics:
A:Introns: 142/3
Query Match 66.7%; Score 34; DB 2; Length 241;
Best Local Similarity 75.0%; Pred. No. 32;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 2 ANHSGNAS 9
Db 64 ANHPGNAT 71

RESULT 10
S36838
acetyl-CoA C-acyltransferase (EC 2.3.1.16), peroxisomal - yeast (Yarrowia lipolytica)
N:Alternate names: 3-oxoacyl-CoA thiolase
C:Species: Yarrowia lipolytica, Candida lipolytica
C:Date: 22-Jan-1994 #sequence_revision 23-Feb-1996 #text_change 09-Jul-2004
C:Accession: S36838; S31331
R:Berninger, G.; Schmidtchen, R.; Casel, G.; Knoerr, A.; Rautenstrauss, K.; Kunau, W.H.;
Eur. J. Biochem. 216, 607-613, 1993
A:Title: Structure and metabolic control of the Yarrowia lipolytica peroxisomal 3-oxoacyl
A:Reference number: S36838; MUID:93387313; PMID:7916689
A:Accession: S36838
A:Molecule type: DNA
A:Residues: 1-414 <BER>
A:Cross-references: UNIPROT:Q05493; EMBL:X69988; NID:g5531; PIDN:CAA49605.1; PID:g5532
C:Superfamily: acetyl-CoA acetyltransferase
C:Keywords: acyltransferase; coenzyme A; peroxisome
Query Match 66.7%; Score 34; DB 2; Length 414;
Best Local Similarity 85.7%; Pred. No. 57;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 4 HSGNASQ 10
Db 264 HAGNASQ 270

RESULT 11
AC0976
probable autotransporter sapB [imported] - Salmonella enterica subsp. enterica serovar Typhi
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AC0976
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AC0976
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1107 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD03303.1; PID:g16504923; GSPDB:GN00176
C:Genetics:
A:Gene: sapB
Query Match 66.7%; Score 34; DB 2; Length 1107;
Best Local Similarity 60.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 1 SANHSGNASQ 10
Db 437 SANHGGSTSK 446

RESULT 12
S41794
SEC3 protein - yeast (Saccharomyces cerevisiae)

```

N/Alternate names: protein YER008c; PSL1 protein
 C/Species: Saccharomyces cerevisiae
 C/Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 09-Jul-2004
 C/Accession: S41794; S50466; S72238
 R/Haarer, B.K.; Petzold, A.S.; Brown, S.S.
 submitted to the EMBL Data Library, July 1993
 A/Description: Identification of mutations that are synthetically lethal with altered yeast
 A/Reference number: S41793
 A/Accession: S41794
 A/Molecule type: DNA
 A/Residues: 1-1336 <HAA>
 A/Cross-references: UNIPROT:P33332; EMBL:L22204; NID:g347714; PID:g347716
 R/Dietrich, F.S.
 submitted to the EMBL Data Library, December 1994
 A/Description: The sequence of S. cerevisiae cosmid 9537, 9581, 9495, 9867, and lambda
 A/Reference number: S50459
 A/Accession: S50466
 A/Molecule type: DNA
 A/Residues: 1-1336 <DIE>
 A/Cross-references: EMBL:U18778; NID:g603592; PID:g603600; MIPS:YER008c
 R/Haarer, B.K.; Corbett, A.; Kweon, Y.; Petzold, A.S.; Silver, P.; Brown, S.S.
 Genetics 144, 495-510, 1996
 A/Title: SEC3 mutations are synthetically lethal with profilin mutations and cause defec
 A/Reference number: S72237; MUID:97044444; PMID:8889515
 A/Accession: S72238
 A/Molecule type: DNA
 A/Residues: 1-1336 <HAW>
 A/Cross-references: EMBL:L22204; NID:g347714; PIDN:AA849380.1; PID:g347716
 C/Genetics:
 A/Gene: SGD:SEC3; PSL1
 A/Cross-references: MIPS:YER008c; SGD:S0000810
 A/Map position: 5R

Query Match 66.7%; Score 34; DB 2; Length 1336;
 Best Local Similarity 70.0%; Pred. No. 2e+02;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 SANHSGNASQ 10
 ||:|||||
 Db 35 SAHHSRNVSQ 44

RESULT 13
 RGYC5
 N/Alternate names: protein CDC25 - yeast (Saccharomyces cerevisiae)
 C/Species: Saccharomyces cerevisiae
 C/Date: 31-Mar-1988 #sequence_revision 31-Mar-1993 #text_change 09-Jul-2004
 C/Accession: A26596; S51442; A23444; S43051; S47990
 R/Broek, D.; Toda, T.; Michaeli, T.; Levin, L.; Birchmeier, C.; Zoller, M.; Powers, S.;
 Cell 48, 789-799, 1987
 A/Title: The S. cerevisiae CDC25 gene product regulates the RAS/adenylate cyclase pathwa
 A/Reference number: A26596; MUID:87131091; PMID:3545497
 A/Accession: A26596
 A/Molecule type: DNA
 A/Residues: 1-1589 <BRO>
 A/Cross-references: UNIPROT:P04821; EMBL:M15458; NID:g171184; PIDN:AAA34478.1; PID:g1711
 R/Fauley, A.
 submitted to the EMBL Data Library, November 1994
 A/Description: The sequence of S. cerevisiae cosmid L2142.
 A/Reference number: S51437
 A/Accession: S51442
 A/Molecule type: DNA
 A/Residues: 1-1589 <PAU>
 A/Cross-references: EMBL:U17247; NID:g577216; PIDN:AB67360.1; PID:g577222; GSPDB:GN0001
 R/Camonis, J.H.; Kalekine, M.; Gondre, B.; Garreau, H.; Boy-Marcotte, E.; Jacquet, M.
 EMO J. 5, 375-380, 1986
 A/Title: Characterization, cloning and sequence analysis of the CDC25 gene which control
 A/Reference number: A23444; MUID:86220116; PMID:3011405
 A/Accession: A23444
 A/Molecule type: DNA
 A/Residues: 1-496, 'y', 498-953, 'LSVIMNLSR', 964-1589 <CAM>
 A/Cross-references: EMBL:X03579; NID:g3483; PIDN:CAA27259.1; PID:g3484

R/Daniel, J.H.
 Curr. Genet. 10, 879-885, 1986
 A/Title: The CDC25 "Start" gene of Saccharomyces cerevisiae: sequencing of the active C-
 A/Reference number: S43051; MUID:88194639; PMID:3329037
 A/Accession: S43051
 A/Molecule type: DNA
 A/Residues: 877-1589 <DAN>
 A/Cross-references: EMBL:X03579
 C/Genetics:
 A/Gene: SGD:CD25; CTN1; MIPS:YLR310c
 A/Cross-references: SGD:S0004301; MIPS:YLR310c
 A/Map position: 12R
 C/Function:
 A/Description: positive control of level of cellular CAMP at the stage at which the cell
 C/Supfamily: budding yeast CD25; CD25-type guanine nucleotide exchange activator hom
 C/Keywords: cell cycle control; transmembrane protein
 F:65-123/Domain: SH3 homology <SH3>
 F:1301-1542/Domain: CDC25-type guanine nucleotide exchange activator homology <SOS>

Query Match 66.7%; Score 34; DB 1; Length 1589;
 Best Local Similarity 70.0%; Pred. No. 2.4e+02;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 SANHSGNASQ 10
 ||:|||||
 Db 13 SAREAGNASQ 22

RESULT 14
 S72384
 hypothetical protein 8 precursor - Enterococcus faecalis plasmid pAD1
 C/Species: Enterococcus faecalis
 C/Date: 12-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 09-Jul-2004
 C/Accession: S72384
 R/Hirt, H.; Wirth, R.; Muscholl, A.
 Mol. Gen. Genet. 252, 640-647, 1996
 A/Title: Comparative analysis of 18 sex pheromone plasmids from Enterococcus faecalis: de
 A/Reference number: S72375; MUID:97074879; PMID:8917306
 A/Accession: S72384
 A/Status: nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-286 <HIR>
 A/Cross-references: UNIPROT:Q47791; EMBL:X96977; NID:g1279406; PIDN:CAA65667.1; PID:g1279
 A/Experimental source: strain OGIX
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1996
 C/Genetics:
 A/Genome: plasmid pAD1
 C/Supfamily: probable pheromone-responsive protein
 F:1-26/Domain: signal sequence #status predicted <SIG>
 F:27-286/Product: hypothetical protein 8 #status predicted <MAT>

Query Match 64.7%; Score 33; DB 2; Length 286;
 Best Local Similarity 62.5%; Pred. No. 61;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 NHSGNASQ 10
 ||:|||||
 Db 170 NHAGNGTQ 177

RESULT 15
 B47172
 ficollin-beta - pig
 C/Species: Sus scrofa domestica (domestic pig)
 C/Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
 C/Accession: B47172
 R/Ichijo, H.; Hellman, U.; Wernstedt, C.; Gonsz, L.J.; Claesson-Welsh, L.; Heldin, C.H.;
 J. Biol. Chem. 268, 14505-14513, 1993
 A/Title: Molecular cloning and characterization of ficollin, a multimeric protein with fit
 A/Reference number: A47172; MUID:93300852; PMID:7686157
 A/Accession: B47172
 A/Status: preliminary
 A/Molecule type: mRNA; protein

A;Residues: 1-326 <ICH>
A;Cross-references: UNIPROT:Q29042; GB:L12345; NID:9294219; PIDN:AAC69641.1; PID:gi22892
A;Experimental source: uterus
A;Note: sequence extracted from NCBI backbone (NCBIN:134468, NCBIP:134470)
F;115-326/Domain: fibrinogen beta/gamma homology <FBG>

Query Match 64.7%; Score 33; DB 2; Length 326;
Best Local Similarity 62.5%; Pred. No. 70;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 NHSGNASQ 10
||| | | | : :
Db 265 NHSGNCAE 272

Search completed: July 18, 2005, 13:41:53
Job time : 16.4 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2005, 12:59:21 ; Search time 75.8 Seconds
(without alignments)
67.557 Million cell updates/sec

Title: SEQ5

Perfect score: 51

Sequence: 1 sanhsnsgaq 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot 03:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	40	78.4	635	Q6LTN6	Q6ltln6 photobacter
2	39	76.5	1461	Q8ZL64	Q8ztl64 salmonella
3	36	70.6	215	Q6EU34	Q6eu34 oryza sativ
4	36	70.6	283	1 TONE NEIGO	O06432 neisseria g
5	36	70.6	298	Q66501	Q66501 aquifex aeo
6	36	70.6	319	Q8H8D6	Q8h8d6 oryza sativ
7	36	70.6	1374	Q9VSU0	Q9vsu0 drosophila
8	36	70.6	1449	Q9U1I2	Q9u1i2 drosophila
9	36	70.6	1450	Q8IQB8	Q8iqb8 drosophila
10	36	70.6	1462	Q9UII3	Q9uii3 drosophila
11	36	70.6	2409	Q960G6	Q960g6 drosophila
12	36	70.6	2786	Q9VSU2	Q9vsu2 drosophila
13	35	68.6	105	Q7DNC1	Q7dnc1 oryza sativ
14	35	68.6	341	1 Alys BPA51	Q38653 bacterioph
15	35	68.6	356	2 Q7YTB3	Q7ytb3 saccoglossu
16	35	68.6	396	2 Q6CNH5	Q6cnh5 kluyveromyc
17	35	68.6	414	2 Q99FX6	Q99fx6 simian rota
18	35	68.6	416	2 Q6RKG7	Q6rkg7 gibberella
19	35	68.6	418	2 Q53183	Q53183 mycobacteri
20	35	68.6	418	2 Q7TYJ4	Q7tyj4 mycobacteri
21	35	68.6	479	2 Q99FX4	Q99fx4 simian rota
22	35	68.6	483	2 Q9PWM8	Q9pwm8 gallus gall
23	35	68.6	495	1 VN53 ROTSL	P15687 simian l1 r
24	35	68.6	495	1 VN53 ROTSP	P35425 simian l1 r
25	35	68.6	496	2 Q99FX5	Q99fx5 simian rota
26	35	68.6	496	2 Q99FX7	Q99fx7 simian rota
27	35	68.6	568	2 Q8P7L4	Q8p7l4 xanthomonas
28	35	68.6	568	2 Q8PIY6	Q8piy6 xanthomonas
29	35	68.6	634	1 HNFA CHICK	Q90867 gallus gall
30	35	68.6	662	2 Q9YX8	Q9yx8 gallus gall
31	35	68.6	752	2 Q7QL98	Q7ql98 anopheles g

32	35	68.6	753	2 Q7PTN5	Q7ptn5 anopheles g
33	35	68.6	761	2 Q7PEX3	Q7pex3 anopheles g
34	35	68.6	820	2 Q754J6	Q754j6 ashbya goss
35	35	68.6	831	2 Q9N9J8	Q9n9j8 leishmania
36	35	68.6	846	2 Q8RPV1	Q8rpv1 streptococc
37	35	68.6	862	2 Q9S7Q2	Q9s7q2 arabidopsis
38	35	68.6	863	2 Q8JBE4	Q8jbe4 human immun
39	35	68.6	1077	2 Q9U209	Q9u209 caenorhabdi
40	35	68.6	1778	2 Q8FCB2	Q8fcb2 escherichia
41	35	68.6	2039	2 Q7S3G9	Q7s3g9 neurospora
42	34	66.7	176	2 Q7PH59	Q7ph59 anopheles g
43	34	66.7	179	1 YC61 SYN3	Y73801 synecocyst
44	34	66.7	207	2 Q7TU22	Q7tu22 prochloroco
45	34	66.7	219	2 Q6KZ76	Q6kz76 picophilus

ALIGNMENTS

RESULT 1

Q6LTN6 PRELIMINARY; PRT; 635 AA.
AC Q6LTN6;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical polar flagellar hook-length control protein Flik.
GN OrderedLocusNames=PBPA0928;
OS Photobacterium profundum (Photobacterium sp. (strain S9)).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Photobacterium.
OX NCBI_TaxID=74109;
RN [1]
RP SEQUENCE FROM N.A.
RA Vezzi A., Campanaro S., D'Angelo M., Simonato F., Vitulo N., Lauro F.,
RA Cestaro A., Malacrida G., Samionati B., Cannata N., Bartlett D.,
RA Valle G.;
RT "Genome analysis of Photobacterium profundum reveals the complexity of
RT high pressure adaptations.";
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; CR378665; CAG19339.1;
DR GO; GO:0009424; C:flagellar hook (sensu Bacteria); IEA.
DR GO; GO:0019861; C:flagellum; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR GO; GO:0009296; P:flagellum biogenesis; IEA.
DR InterPro; IPR001635; Flag hook.
DR Pfam; PF02120; Flg hook; I.
KW Complete proteome; Flagellum; Hypothetical protein.
SQ SEQUENCE 635 AA; 65949 MW; 3D5E7E78F8C44171 CRC64;

Query Match 78.4%; Score 40; DB 2; Length 635;
Best Local Similarity 70.0%; Pred. No. 26;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 SANHSGNASQ 10
| | | | | : |
DB 594 SGNHSGNAGQ 603

RESULT 2

Q8ZL64 PRELIMINARY; PRT; 1461 AA.
AC Q8ZL64;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative inner membrane protein.
GN OrderedLocusNames=STM3691;
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]

```

RP SEQUENCE FROM N.A.
RC STRAIN=LT2;
RX MEDLINE=21534948; PubMed=11677609; DOI=10.1038/35101614;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2."
RL Nature 413:852-856(2001).
DR EMBL; AS008871; AAL22550.1; -.
DR GO; GO:0019867; C:outer membrane; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR Pfam; PF05658; Hep Hag; 10.
DR Pfam; PF05662; HIM; 13.
DR Pfam; PF03895; Yada; 1.
KW Complete proteome.
SQ SEQUENCE 1461 AA; 147836 MW; 91C59A87E7282254 CRC64;

Query Match 76.5%; Score 39; DB 2; Length 1461;
Best Local Similarity 70.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 SANHSGNASQ 10
Db 328 SASHTGNASQ 337
:::|||||

RESULT 3
Q6EU34 PRELIMINARY; PRT; 215 AA.
AC Q6EU34;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein OJ1123_G04.17.
GN Name=OJ1123_G04.17;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP004178; BAD27836.1; -.
KW Hypothetical protein.
SQ SEQUENCE 215 AA; 22508 MW; D4A6A8EB0AA473CB CRC64;

Query Match 70.6%; Score 36; DB 2; Length 215;
Best Local Similarity 70.0%; Pred. No. 48;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 SANHSGNASQ 10
Db 94 ANYSGKASQ 103
:::|||||

RESULT 4
TONB_NEIGO STANDARD; PRT; 283 AA.
AC O06432;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE TonB protein.
GN Name=tonB;
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;

```

```

RN RP SEQUENCE FROM N.A.
RC STRAIN=PA19;
RX MEDLINE=97285757; PubMed=9140974;
RA Biswas G.D., Anderson J.E., Sparling P.F.;
RT "Cloning and functional characterization of Neisseria gonorrhoeae
tonB, exbB and exbD genes."
RL Mol. Microbiol. 24:169-179(1997).
CC -!- FUNCTION: Pathways of utilization of iron bound to transferrin,
CC lactoferrin and hemoglobin but not to haemin or citrate where
CC dependent on the tonB system.
CC -!- SUBUNIT: The accessory proteins exbB and exbD seem to form a
CC complex with tonB (By similarity).
CC -!- SUBCELLULAR LOCATION: Periplasmic. Anchored to the cytoplasmic
CC membrane via its N-terminal signal-like sequence, spans the
CC periplasm (By similarity).
CC -!- SIMILARITY: Belongs to the tonB family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U79563; AAC45286.1; -.
CC InterPro; IPR003538; TonB.
CC InterPro; IPR006260; TonB_C.
CC Pfam; PF03544; TonB; 1.
CC PRINTS; PR01374; TONBPROTEIN.
CC TIGRFAMs; TIGR01352; tonB Cterm; 1.
CC Inner membrane, periplasmic; Protein transport; Signal-anchor;
CC Transmembrane; Transport.
CC DOMAIN 1 5 Cytoplasmic (Potential).
CC TRANSMEM 6 27 Signal-anchor (Potential).
CC DOMAIN 28 283 Periplasmic (Potential).
CC SEQUENCE 283 AA; 28749 MW; 3CD3F8353B45748 CRC64;

Query Match 70.6%; Score 36; DB 1; Length 283;
Best Local Similarity 87.5%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ANHSGNAS 9
Db 136 AEHSGNAS 143
|||
|||
|||

RESULT 5
O66501 PRELIMINARY; PRT; 298 AA.
AC O66501;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein aq_091.
GN OrderedLocusNames=AQ_091;
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320; DOI=10.1038/32831;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aujay M., Huber R.,
RA Feldman R.A., Short J.M., Olsen G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
aeolicus."
RL Nature 392:353-358(1998).
CC -!- SIMILARITY: Contains 1 sigma-54 factor interaction ATP-binding
CC domain.
CC EMBL; AF000673; AAC06468.1; -.

```

```

DR PIR: B70309; B70309.
DR GO: GO:0003677; F:DNA binding; IEA.
DR GO: GO:0004871; F:signal transducer activity; IEA.
DR GO: GO:0000160; P:two-component signal transduction system (p. . .; IEA.
DR InterPro: IPR000014; PAS.
DR InterPro: IPR002078; Sig54_interact.
DR Pfam: PF00989; PAS; 2
DR Pfam: PF00158; Sigma54_activat; 1.
DR SMART; SM00091; PAS; 2.
DR PROSITE; PS00675; SIGMAS4_INTERACT_1; 1.
DR PROSITE; PS00045; SIGMAS4_INTERACT_4; 1.
KW ATP-binding; Complete proteome; DNA-binding; Transcription;
KW Transcription regulation.
SQ SEQUENCE 298 AA; 33514 MW; 6C4389D727CDAC29 CRC64;

Query Match 70.6%; Score 36; DB 2; Length 298;
Best Local Similarity 75.0%; Pred. No. 70;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 NHSGNASQ 10
DB 146 NHNGNASK 153

RESULT 6
Q8H8D6 PRELIMINARY; PRT; 319 AA.
AC Q8H8D6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Putative glutathione S-transferase.
GN Names=OJ1006F06.9;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA Wing R.A., Yu Y., Soderlund C., Kim H.-R., Rambo T., Saski C.,
RA Currie J., Collura K.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC099399; AA05499.1; -.
DR HSSP; Q92P62; 1A09.
DR Gramene; Q8H8D6; -.
DR GO: GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR004046; GST_Cterm.
DR InterPro; IPR010987; GST_C-like.
DR Pfam; PF00043; GST_C; 1.
DR Pfam; PF02798; GST_N; 1.
KW Transferase.
SQ SEQUENCE 319 AA; 35704 MW; 0D4CB62D35C6D1D8 CRC64;

Query Match 70.6%; Score 36; DB 2; Length 319;
Best Local Similarity 70.0%; Pred. No. 76;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 SANHSGNASQ 10
DB 274 SAAHAGNAQ 283

RESULT 7
Q9VSU0 PRELIMINARY; PRT; 1374 AA.
AC Q9VSU0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE C54821-PB.
GN Name=Requila; ORFNames=CG4821;

```

```

OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA April J.F., Agbayani A., An H.J., Andrews-Pfannkuch C., Baldwin D.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodgett W.K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila
melanogaster euchromatic genome sequence."
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
a genomics perspective."
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,

```

RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.O.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review";
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN [5]
RN SEQUENCE FROM N.A.
RG FlyBase;
RG Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [6]
RN SEQUENCE FROM N.A.
RG FlyBase;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to peptidase family 51.
DR EMBL; AF003553; AAF50321.2; -.
DR HSSP; P00760; 1EZX.
DR FlyBase; FBgn0023479; Tequila.
DR GO; GO:0004295; F:trypsin activity; NAS.
DR GO; GO:0006508; P:proteolysis and peptidolysis; NAS.
DR InterPro; IPR000194; Atpase_a/bcentre.
DR InterPro; IPR002557; Chitin_bind_PeR.
DR InterPro; IPR002172; Ldl_receptor_A.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR InterPro; IPR001190; Sscr_receptor.
DR Pfam; PF01607; CEM_14; 1.
DR Pfam; PF00057; Ldl_recept_a; 2.
DR Pfam; PF00530; SRCR_2.
DR Pfam; PF00089; Trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR PRINTS; PR00258; SPERACTRCPTR.
DR SMART; SM00494; ChtBD2; 1.
DR SMART; SM00192; LDLa; 2.
DR SMART; SM00202; SR; 2.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00152; ATPASE_ALPHA_BETA; UNKNOWN_1.
DR PROSITE; PS00530; SRCR; 2.
DR Pfam; PF00089; Trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR PRINTS; PR00258; SPERACTRCPTR.
DR SMART; SM00494; ChtBD2; 1.
DR SMART; SM00192; LDLa; 2.
DR SMART; SM00202; SR; 2.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00152; ATPASE_ALPHA_BETA; UNKNOWN_1.
DR PROSITE; PS01209; LDLRA_1; 1.
DR PROSITE; PS00668; LDLRA_2; 2.
DR PROSITE; PS00420; SRCR_1; 2.
DR PROSITE; PS0287; SRCR_2; 2.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 1374 AA; 153015 MW; D9129FD9360BED91 CRC64;

Query Match 70.6%; Score 36; DB 2; Length 1374;
Best Local Similarity 75.0%; Pred. No. 4.1e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 3 NHSGNASQ 10
Db 561 NHSGNAQE 568

RESULT 8

ID Q9U112 PRELIMINARY; PRT; 1449 AA.
AC Q9U112;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE GRAAL protein.
GN Name=Tequila; Synonyms=graal;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;
RN [1]
RN SEQUENCE FROM N.A.
RP Munier A.I., Medzhitov R., Janeway C., Hoffmann J.A., Lagueux M.;
RA Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to peptidase family 51.
DR EMBL; AJ251803; CAB64653.1; -.
DR HSSP; P00760; 1EZX.
DR FlyBase; FBgn0023479; Tequila.
DR GO; GO:0004295; F:trypsin activity; NAS.
DR GO; GO:0006508; P:proteolysis and peptidolysis; NAS.
DR InterPro; IPR000194; Atpase_a/bcentre.
DR InterPro; IPR002557; Chitin_bind_PeR.
DR InterPro; IPR002172; Ldl_receptor_A.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR InterPro; IPR001190; Sscr_receptor.
DR Pfam; PF01607; CEM_14; 2.
DR Pfam; PF00057; Ldl_recept_a; 2.
DR Pfam; PF00530; SRCR_2.
DR Pfam; PF00089; Trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR PRINTS; PR00258; SPERACTRCPTR.
DR SMART; SM00494; ChtBD2; 2.
DR SMART; SM00192; LDLa; 2.
DR SMART; SM00202; SR; 2.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00152; ATPASE_ALPHA_BETA; UNKNOWN_1.
DR PROSITE; PS00940; CHIT_BIND_II; 2.
DR PROSITE; PS01209; LDLRA_1; 1.
DR PROSITE; PS00668; LDLRA_2; 2.
DR PROSITE; PS00420; SRCR_1; 2.
DR PROSITE; PS0287; SRCR_2; 2.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 1449 AA; 160045 MW; F3BC906543CAB6ED CRC64;

Query Match 70.6%; Score 36; DB 2; Length 1449;
Best Local Similarity 75.0%; Pred. No. 4.3e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 3 NHSGNASQ 10
Db 636 NHSGNAQE 643

RESULT 9

OY Q8IQB8 PRELIMINARY; PRT; 1450 AA.
AC Q8IQB8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE CG4821-PD.
GN Name=Tequila; ORFNames=CG4821;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,

RA Abril J.P., Agbayani A., An H.J., Andrews-Pfannkuch C., Baldwin D.,
 RA Ballaw R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Bouchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foslér C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodagef, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195 (2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426065; PubMed=12537568;
 RA Celiniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
 RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
 RA George R.A., Hoskins R.A., Leverty T., Muzny D.M., Nelson C.R.,
 RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
 RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
 RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
 RT "Finishing a whole-genome shotgun: Release 3 of the *Drosophila*
 RT *melanogaster* euchromatic genome sequence.";
 RL Genome Biol. 3:RESEARCH0079-RESEARCH0079 (2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426070; PubMed=12537573;
 RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
 RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
 RA Ashburner M., Celiniker S.E.;
 RT "The transposable elements of the *Drosophila melanogaster* euchromatin:
 RT a genomics perspective.";
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084 (2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426069; PubMed=12537572;
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Bettencourt B.R., Celiniker S.E., de Grey A.D., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a
 RT systematic review.";
 RL Genome Biol. 3:RESEARCH0083-RESEARCH0083 (2002).
 RN [5]
 RP SEQUENCE FROM N.A.
 RG FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RG FlyBase;

RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: Belongs to peptidase family S1.
 DR EMBL; AB003553; AAN11984.1; -;
 DR HSSP; P00760; 1EZX.
 DR IntAct; Q8IQB8; -;
 DR FlyBase; FBgn0023479; Tequila.
 DR GO; GO:0004295; F:trypsin activity; NAS.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; NAS.
 DR InterPro; IPR000194; ATPase_a/bcentre.
 DR InterPro; IPR002557; Chitin_bind_Pera.
 DR InterPro; IPR002172; LDL receptor_A.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001314; Peptidase_S1A.
 DR InterPro; IPR009003; Pept_Ser_Cys.
 DR InterPro; IPR001190; Srrt_receptor.
 DR Pfam; PF01607; CBM_14; 2.
 DR Pfam; PF00530; SRRG; 2.
 DR Pfam; PF00089; Trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR PRINTS; PR00261; LDLRECEPTOR.
 DR PRINTS; PR00258; SPERACTRCPTR.
 DR SMART; SM00494; ChtBD2; 2.
 DR SMART; SM00192; LDLa; 2.
 DR SMART; SM00202; SR; 2.
 DR SMART; SM00020; Tryp_Spc; 1.
 DR PROSITE; PS00152; ATPASE_ALPHA_BETA; UNKNOWN_1.
 DR PROSITE; PS0940; CHIT_BIND_II; 2.
 DR PROSITE; PS01209; LDLRA_1; 1.
 DR PROSITE; PS50068; LDLRA_2; 2.
 DR PROSITE; PS00420; SRCH_1; 2.
 DR PROSITE; PS00287; SRCH_2; 2.
 DR PROSITE; PS0240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydrolase; Protease; Serine protease.
 SQ SEQUENCE 1450 AA; 160130 MW; E4CC56D104CC3735 CRC64;
 QY 3 NHSGNASQ 10
 DB 637 NHSGNAQE 644
 Query Match 70.6%; Score 36; DB 2; Length 1450;
 Best Local Similarity 75.0%; Pred.No. 4.3e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 RESULT 10
 Q9ULI3 PRELIMINARY; PRT; 1462 AA.
 ID Q9ULI3
 AC Q9ULI3
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE GRAAL protein.
 GN Names=Tequila; Synonyms=graal;
 OS *Drosophila melanogaster* (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; *Drosophila*.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Munier A.I., Medzhitov R., Janeway C., Hoffmann J.A., Lagueux M.;
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: Belongs to peptidase family S1.
 DR EMBL; AJ251802; CAB64652.1; -;
 DR HSSP; P00760; 1EZX.
 DR FlyBase; FBgn0023479; Tequila.
 DR GO; GO:0004295; F:trypsin activity; NAS.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; NAS.
 DR InterPro; IPR000194; ATPase_a/bcentre.
 DR InterPro; IPR002557; Chitin_bind_Pera.

```
DR InterPro: IPR002172; LDL receptor A.
DR InterPro: IPR001254; Peptidase S1_
DR InterPro: IPR001314; Peptidase S1A.
DR InterPro: IPR009003; Pept_Ser_Cys.
DR InterPro: IPR001190; Srcr_receptor.
DR Pfam: PF01607; CBM_14; 2.
DR Pfam: PF00057; Ldl_recept_a; 2.
DR Pfam: PF00530; SRCR; 2.
DR Pfam: PF00089; Trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR PRINTS: PR00261; LDLRECEPTOR.
DR PRINTS: PR00258; SPERACTRCPTR.
DR SMART: SM00494; ChtBD2; 2.
DR SMART: SM00192; LDLA; 2.
DR SMART: SM00202; SR; 2.
DR SMART: SM00089; Trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR PRINTS: PR00261; LDLRECEPTOR.
DR PRINTS: PR00258; SPERACTRCPTR.
DR SMART: SM00494; ChtBD2; 2.
DR SMART: SM00192; LDLA; 2.
DR SMART: SM00202; SR; 2.
DR SMART: SM00089; Trypsin; 1.
DR PROSITE: PS00152; ATPEASE ALPHA BETA; UNKNOWN_1.
DR PROSITE: PS0940; CHIT BIND II; 8.
DR PROSITE: PS01209; LDLRA_1; 1.
DR PROSITE: PS00668; LDLRA_2; 2.
DR PROSITE: PS00420; SRCR_1; 2.
DR PROSITE: PS0287; SRCR_2; 2.
DR PROSITE: PS0240; TRYPSIN DOM; 1.
DR PROSITE: PS0134; TRYPSIN HIS; UNKNOWN_1.
DR PROSITE: PS00135; TRYPSIN SER; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 1462 AA; 161444 MW; E3AC494FE140F37A CRC64;

Query Match 70.6%; Score 36; DB 2; Length 1462;
Best Local Similarity 75.0%; Pred. No. 4.4e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 NHSGNASQ 10
DB 649 NHSGNAQE 656

RESULT 11
Q960G6 PRELIMINARY; PRT; 2409 AA.
AC Q960G6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE SD0286OP.
GN Name=Tequila; Synonym=CG4821;
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Paeleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RA Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to peptidase family S1.
DR EMBL: AY052072; AAK93496.1; -.
DR HSP: P00760; 1EZX
DR FlyBase: FBgn023479; Tequila.
DR GO: GO:0044295; F:trypsin activity; NAS.
DR GO: GO:006508; P:proteolysis and peptidolysis; NAS.
DR InterPro: IPR001194; ATPase a/bcentre.
DR InterPro: IPR002557; Chitin bind PerA.
DR InterPro: IPR002172; LDL receptor A.
DR InterPro: IPR001254; Peptidase S1_
DR InterPro: IPR001314; Peptidase S1A.
DR InterPro: IPR009003; Pept_Ser_Cys.
DR Pfam: PF01607; CBM_14; 10.

DR InterPro: IPR002172; LDL receptor A.
DR InterPro: IPR001254; Peptidase S1_
DR InterPro: IPR001314; Peptidase S1A.
DR InterPro: IPR009003; Pept_Ser_Cys.
DR Pfam: PF01607; CBM_14; 2.
DR Pfam: PF00057; Ldl_recept_a; 2.
DR Pfam: PF00530; SRCR; 2.
DR Pfam: PF00089; Trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR PRINTS: PR00261; LDLRECEPTOR.
DR PRINTS: PR00258; SPERACTRCPTR.
DR SMART: SM00494; ChtBD2; 2.
DR SMART: SM00192; LDLA; 2.
DR SMART: SM00202; SR; 2.
DR SMART: SM00089; Trypsin; 1.
DR PROSITE: PS00152; ATPEASE ALPHA BETA; UNKNOWN_1.
DR PROSITE: PS0940; CHIT BIND II; 8.
DR PROSITE: PS01209; LDLRA_1; 1.
DR PROSITE: PS00668; LDLRA_2; 2.
DR PROSITE: PS00420; SRCR_1; 2.
DR PROSITE: PS0287; SRCR_2; 2.
DR PROSITE: PS0240; TRYPSIN DOM; 1.
DR PROSITE: PS0134; TRYPSIN HIS; UNKNOWN_1.
DR PROSITE: PS00135; TRYPSIN SER; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 1462 AA; 161444 MW; E3AC494FE140F37A CRC64;

Query Match 70.6%; Score 36; DB 2; Length 2409;
Best Local Similarity 75.0%; Pred. No. 7.7e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 NHSGNASQ 10
DB 1596 NHSGNAQE 1603

RESULT 12
Q9VSU2 PRELIMINARY; PRT; 2786 AA.
AC Q9VSU2; Q9VSU1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE CG4821-PA.
GN Name=Tequila; ORFNames=CG4821;
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burkis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
```

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M., Palazolo M., Pittman K.A., Pan S.S., Pollard J., Puri V., Reese M.G., Reinert K., Remington K.S., Saunders R.D., Scheeler F., Shen H., Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissenbach J., Williams S.M., Woodgett, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zhou L., Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O., Svirkas R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RA "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 RN [2]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=22426065; PubMed=12537568;
 RA Celnikier S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A., Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A., George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R., Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J., Svirkas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C., Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
 RA "Finishing a whole-genome shotgun: Release 3 of the *Drosophila melanogaster* euchromatic genome sequence.";
 RT Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
 RN [3]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=22426070; PubMed=12537573;
 RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirkas R., Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M., Ashburner M., Celnikier S.E.;
 RA "The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective.";
 RT Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
 RN [4]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=22426069; PubMed=12537572;
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S., Hradscky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E., Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P., Bettencourt B.R., Celnikier S.E., de Grey A.D., Drysdale R.A., Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q., Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M., Lewis S.E.;
 RA "Annotation of the *Drosophila melanogaster* euchromatic genome: a systematic review.";
 RT Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
 RN [5]
 RN SEQUENCE FROM N.A.
 RP FlyBase;
 RG Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RN SEQUENCE FROM N.A.
 RP FlyBase;
 RG Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: Belongs to peptidase family 51.
 DR EMBL; AE003553; AAF50319.3; -;
 DR HSSP; P00760; IEZX.
 DR FlyBase; FBgn023479; Tegula.
 DR GO; GO:0004295; F.trypsin activity; NAS.
 DR GO; GO:0006508; P.proteolysis and peptidolysis; NAS.
 DR InterPro; IPR000194; Arpase_a/bcentre.
 DR InterPro; IPR002557; Chitin bind Pera.
 DR InterPro; IPR002172; LDL receptor A.
 DR InterPro; IPR001254; Peptidase S1.
 DR InterPro; IPR001314; Peptidase_S1A.
 DR InterPro; IPR009003; Pept_Ser_Cys.
 DR InterPro; IPR001190; Srcr_receptor.
 DR Pfam; PF01607; CBM 14; 15.
 DR Pfam; PF00057; Ldl_recept_a; 2.
 DR Pfam; PF00530; SRCR; 2.
 DR Pfam; PF00089; Trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.

DR PRINTS; PR00261; LDLRECEPTOR.
 DR PRINTS; PR00258; SPERACTRCPTR.
 DR SMART; SM00494; ChtBD2; 15.
 DR SMART; SM00192; LDLa; 2.
 DR SMART; SM00202; SR; 2.
 DR SMART; SM00020; TYP SPC; 1.
 DR PROSITE; PS00152; ATPASE ALPHA BETA; UNKNOWN_1.
 DR PROSITE; PS00940; CHIT BIND II; 11.
 DR PROSITE; PS01209; LDLRA_1; 1.
 DR PROSITE; PS00068; LDLRA_2; 2.
 DR PROSITE; PS00420; SRCR 1; 2.
 DR PROSITE; PS0287; SRCR 2; 2.
 DR PROSITE; PS0240; TRYPSIN DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydrolase; Protease; Serine protease.
 SQ SEQUENCE 2786 AA; 308363 MW; 2EED7A7DA5002C76 CRC64;
 Query Match 70.6%; Score 36; DB 2; Length 2786;
 Best Local Similarity 75.0%; Pred. No. 9.1e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 3 NHSGNASQ 10
 Db 1973 NHSGNAQE 1980
 RESULT 13
 Q7DNC1 PRELIMINARY; PRT; 105 AA.
 AC Q7DNC1
 DT 05-JUL-2004 (TREMBlrel. 27, Created)
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
 DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
 DE OSJNB0046K02.2 protein.
 GN Name=OSJNB0046K02.2;
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OX NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX PubMed=12447439; DOI=10.1038/nature01183;
 RA Feng Q., Zhang Y., Hao P., Wang S., Fu G., Huang Y., Li Y., Zhu J., Liu Y., Hu X., Jia P., Zhang Y., Zhao Q., Ying K., Yu S., Tang Y., Weng Q., Zhang L., Lu Y., Mu J., Lu Y., Zhang L.S., Yu Z., Fan D., Liu X., Lu T., Li C., Wu Y., Sun T., Lei H., Li T., Hu H., Guan J., Wu M., Zhang R., Zhou B., Chen Z., Chen L., Jin Z., Wang R., Yin H., Cai Z., Ren S., Lv G., Gu W., Zhu G., Tu Y., Jia J., Zhang Y., Chen J., Kang H., Chen X., Shao C., Sun Y., Hu Q., Zhang X., Zhang W., Wang L., Ding C., Sheng H., Gu J., Chen S., Ni L., Zhu F., Chen W., Lan L., Lai Y., Cheng Z., Gu M., Jiang J., Li J., Hong G., Xue Y., Han B.;
 RA "Sequence and analysis of rice chromosome 4.";
 RL Nature 420:316-320(2002).
 RN [2]
 RN SEQUENCE FROM N.A.
 RA Han B., Feng Q., Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X., Liu Y.L., Mu J., Yu Z., Chen L., Fan D.L., Weng Q.J., Zhang L., Lu Y.Q., Yu S.L., Liu X.H., Lu T.T., Zhang Y.J., Lu Y., Li C., Li T., Zhang Y., Hu H., Jia P.X., Qian Y.M., Ying K., Zhou B., Chen Z.H., Hao P., Zhang L., Wu M., Zhang R.Q., Guan J.F., Fu G., Wang S.Y., Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F., Jia J., Yin H.F., Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y., Shao C.Y., Sun Y., Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W., Sheng H.H., Gu J.L., Chen S.T., Ni L., Zhu F.H., Hong G.F.;
 RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BX569684; CAE05792.1; -;
 SQ SEQUENCE 105 AA; 11584 MW; 4D8D570F0E20E60F CRC64;
 Query Match 68.6%; Score 35; DB 2; Length 105;
 Best Local Similarity 66.7%; Pred. No. 34;

```
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 2 ANHSGNASQ 10
Db 74 ANHAGHTSQ 82
RESULT 14
ALYS BPA51
ID Alys BPA51 STANDARD; PRT; 341 AA.
AC Q38653;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 05-JUL-2004 (Rel. 44, Last annotation update)
DE Endolysin (EC 3.5.1.28) (N-acetylmuramoyl-L-alanine amidase).
GN Name=PLJ511;
OS Bacteriophage A511.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae.
OX NCBI_TaxID=40523;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96020653; PubMed=8577256;
RA Loessner M.J., Wendlinger G., Scherer S.;
RT "Heterogeneous endolysins in Listeria monocytogenes bacteriophages: a new class of enzymes and evidence for conserved holin genes within the RT siphovital lysis cassettes.";
RL Mol. Microbiol. 16:1231-1241(1995).
CC -!- CATALYTIC ACTIVITY: Hydrolyzes the link between N-acetylmuramoyl residues and L-amino acid residues in certain bacterial cell-wall glycopeptides.
CC -!- DEVELOPMENTAL STAGE: Expressed at about 20 minutes after infection.
CC -!- SIMILARITY: Belongs to the N-acetylmuramoyl-L-alanine amidase family 2.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
-----
EMBL; X85010; CAA59368.1; -.
PIR; S69802; S69802.
MEROPS; M15.950; -.
InterPro; IPR002502; Amidase_2.
Pfam; PF01510; Amidase_2; 1.
SMART; SM00644; Ami_2; 1.
KW Bacteriolytic enzyme; Cell wall; Hydrolase; Repeat.
SQ SEQUENCE 341 AA; 36477 MW; 86C40B73F0FD8547 CRC64;
Query Match 68.6%; Score 35; DB 1; Length 341;
Best Local Similarity 77.8%; Pred. No. 1.3e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 2 ANHSGNASQ 10
Db 231 ANYSGTASQ 239
```

```
RESULT 15
Q7YTB3
ID Q7YTB3 PRELIMINARY; PRT; 356 AA.
AC Q7YTB3;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Brain factor 1 (Fragment).
GN Name=bf-1;
OS Saccoglossus kowalevskii.
OC Eukaryota; Metazoa; Hemichordata; Enteropneusta; Harrimaniidae;
OC Saccoglossus.
```

```
OX NCBI_TaxID=10224;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22721445; PubMed=12837244; DOI=10.1016/S0092-8674(03)00469-0;
RA Lowe C.J., Wu M., Salic A., Evans L., Lander E., Stange-Thomann N., Gruber C.E., Gerhart J., Kirschner M.;
RT "Anterior-posterior patterning in hemichordates and the origins of the RT chordate nervous system.";
RL Cell 113:853-865(2003).
DR EMBL; AY318741; AAP79301.1; -.
DR HSSP; Q99958; LDSV.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; P:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001766; TF_Fork_head.
DR InterPro; IPR009058; Wing_hlx_DNA_bnd.
DR Pfam; PF00250; Fork_head; 1.
DR ProDom; PD000425; TF_Fork_head; 1.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00657; FORK_HEAD_1; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS00039; FORK_HEAD_3; 1.
FT NON_TER 356 356
SQ SEQUENCE 356 AA; 39334 MW; 4050B6891740608D CRC64;
Query Match 68.6%; Score 35; DB 2; Length 356;
Best Local Similarity 85.7%; Pred. No. 1.4e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 SANHSGN 7
Db 32 NANHSGN 38
```

Search completed: July 18, 2005, 13:33:26
Job time : 78.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2005, 12:59:21 ; Search time 84.2 Seconds
(without alignments)
45.934 Million cell updates/sec

Title: SEQ5

Perfect score: 51

Sequence: 1 sanhsnag 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_16Dec04:*

1: Geneseqp19808:*

2: Geneseqp19908:*

3: Geneseqp20008:*

4: Geneseqp20018:*

5: Geneseqp20028:*

6: Geneseqp20038:*

7: Geneseqp20038a:*

8: Geneseqp20048:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	51	100.0	10	3	AAY71090 Synthetic
2	42	82.4	8	3	AAY70988 Peptide 1
3	39	76.5	144	4	Aau64763 Propionib
4	39	76.5	144	6	Abm61282 Propionib
5	39	76.5	1461	6	Abu47415 Protein e
6	36	70.6	283	6	Abp78403 N. gonorr
7	26	70.6	1186	4	Abb60992 Drosophi
8	35	68.6	328	6	Abp78234 N. gonorr
9	35	68.6	341	2	Aar91515 Listeria
10	35	68.6	418	6	Abu34479 Protein e
11	35	68.6	418	6	Abu36792 Protein e
12	35	68.6	762	7	Adc94648 E. faeciu
13	35	68.6	1246	6	Abu1849 Protein e
14	35	68.6	1778	4	Abb52677 Escherich
15	34	66.7	12	6	Abp70279 Amino aci
16	34	66.7	12	6	Abp70280 Amino aci
17	34	66.7	12	6	Abp70306 Amino aci
18	34	66.7	12	6	Abp70281 Amino aci
19	34	66.7	12	6	Abp70282 Amino aci
20	34	66.7	36	2	AAY40015 Peptide s
21	34	66.7	71	4	Aau64931 Propionib
22	34	66.7	71	6	Abm61450 Propionib
23	34	66.7	72	2	AAY12313 Human 5'
24	34	66.7	101	8	Adj66700 hrp patho
25	34	66.7	145	2	AAY40012 Peptide s

26	34	66.7	156	7	ABO80146	ABO80146 Pseudomon
27	34	66.7	198	5	ABB89480	ABB89480 Human pol
28	34	66.7	242	8	ADS24314	ADS24314 Bacterial
29	34	66.7	257	6	ABG99954	ABG99954 Human nov
30	34	66.7	257	7	ADC37399	ADC37399 Nuclear f
31	34	66.7	257	7	ADC37397	ADC37397 Nuclear f
32	34	66.7	293	3	AGC23236	AGC23236 Arabidops
33	34	66.7	323	3	AB432391	AB432391 Human ORF
34	34	66.7	324	3	AAY52394	AAY52394 Human tra
35	34	66.7	324	3	AAY84895	AAY84895 A human p
36	34	66.7	324	4	AAM93835	AAM93835 Human pol
37	34	66.7	324	4	AAM93650	AAM93650 Human pol
38	34	66.7	324	5	AAU85620	AAU85620 Lung anti
39	34	66.7	324	6	ABU69592	ABU69592 Human lun
40	34	66.7	324	6	ABU66495	ABU66495 Lung canc
41	34	66.7	324	7	ADH47445	ADH47445 Human lun
42	34	66.7	324	8	ADL31481	ADL31481 Human pro
43	34	66.7	324	8	ADL31872	ADL31872 Human pro
44	34	66.7	324	8	ADJ21364	ADJ21364 Human lun
45	34	66.7	324	8	ADR73502	ADR73502 Human SLC

ALIGNMENTS

RESULT 1

AAAY71090	ID	AAAY71090 standard; peptide; 10 AA.
XX	AC	AAAY71090;
XX	DT	21-SEP-2000 (first entry)
XX	DE	Synthetic linker peptide #5 encoded by MV10JA oligonucleotide linker.
XX	KW	Llama; HC-V; heavy chain variable domain; antigen binding protein;
XX	KW	linker; conformational flexibility; multivalent binding protein; bi-head;
XX	KW	human chorionic gonadotropin; hCG; immunoassay; agglutination assay;
XX	OS	Synthetic.
XX	PH	Key
XX	FT	Peptide
XX	FT	Location/Qualifiers
XX	FT	2..9
XX	FT	/label= Peptide linker 5
XX	FT	/note= "Planked by one residue from N- and C-termini of
XX	FT	HCV fragment"
XX	PN	WO200024884-A2.
XX	PD	04-MAY-2000.
XX	PF	22-OCT-1999; 99WO-EP008323.
XX	PR	27-OCT-1998; 98WO-EP006991.
XX	PR	22-APR-1999; 99EP-00303118.
XX	PA	(UNIL) UNILEVER PLC.
XX	PA	(UNIL) UNILEVER NV.
XX	PA	(HIND-) HINDUSTAN LEVER LTD.
XX	PI	Frenken LGJ, Howell S, Van Der Vaart JM;
XX	DR	WPI; 2000-350728/30.
XX	DR	N-PSDB; AAD00665.
XX	PT	Use of a linker whose amino acid sequence confers restricted
XX	PT	conformational flexibility to generate multivalent and multispecific
XX	XX	antigen binding proteins.
XX	PS	Example 1.1d; Page 21; 50pp; English.
XX	CC	The present sequence is the synthetic linker peptide #5, encoded by the

CC oligonucleotide linker fragment, MV10JA. It consists of the last residue
CC of the N-terminal HC-V fragment (S) and the first residue of the C-
CC terminal HC-V fragment (Q), intersected by the connecting linker peptide.
CC It is used for the construction of *Saccharomyces cerevisiae* episomal
CC expression plasmid, pUR5334, encoding anti-hCG-anti-RR6 bispecific
CC biheads, containing the linker peptide. The peptide linker confers
CC restricted conformational flexibility for linking binding units in a
CC multivalent binding protein. The linker is used to generate multivalent
CC or multispecific antigen binding proteins for immunoassays, agglutination
CC assays or for purification
XX

SQ Sequence 10 AA;
Query Match 100.0%; Score 51; DB 3; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.0038;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SANHSGNASQ 10
Db 1 SANHSGNASQ 10
|||||

RESULT 2
AAAY70988
ID AAY70988 standard; peptide; 8 AA.
XX
AC AAY70988;
XX
DT 21-SEP-2000 (first entry)
XX
DE Peptide linker-5 for preparation of multivalent antigen binding protein.
XX
KW Linker; multivalent antigen binding protein; conformational flexibility;
KW cell wall protein; CWP; cellobiohydrolase; CBH; immunoglobulin; llama;
KW heavy chain variable domain; HC-V; immunoassay; agglutination assay;
KW Camelid; purification.
XX
OS Synthetic.
XX
FN WO200024884-A2.
XX
PD 04-MAY-2000.
XX
PF 22-OCT-1999; 99WO-EP008323.
XX
PR 27-OCT-1998; 98WO-EP006991.
PR 22-APR-1999; 99EP-00303118.
XX
PA (UNIL) UNILEVER PLC.
PA (UNIL) UNILEVER NV.
PA (HIND-) HINDUSTAN LEVER LTD.
XX
PI Frenken LGJ, Howell S, Van Der Vaart JM;
XX
DR WPI; 2000-350728/30.
XX
PT Use of a linker whose amino acid sequence confers restricted
PT conformational flexibility to generate multivalent and multispecific
PT antigen binding proteins.
XX
PS Disclosure; Page 6; 50pp; English.
XX
CC The present sequence is a peptide linker molecule, that confers
CC restricted conformational flexibility and is used for linking several
CC antigen binding units in a multivalent or multispecific protein. The
CC peptide linker groups are derived from naturally occurring cell wall
CC proteins (CWP) or cellobiohydrolases (CBH), e.g., CWP1 or CBH1P. The
CC antigen binding unit, consists of heavy chain variable domain (HC-V),
CC derived from an immunoglobulin naturally devoid of light chains, e.g.,
CC from camelids. The linker is used to generate multivalent or
CC multispecific antigen binding proteins for immunoassays, agglutination
CC assays or for purification
XX

SQ Sequence 8 AA;
Query Match 82.4%; Score 42; DB 3; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.8e+06;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 ANHSGNAS 9
Db 1 ANHSGNAS 8
|||||

RESULT 3
AAU64763
ID AAU64763 standard; protein; 144 AA.
XX
AC AAU64763;
XX
DT 27-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #25659.
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
FN WO200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US012865.
XX
PR 21-APR-2000; 2000US-0199047P.
PR 02-JUN-2000; 2000US-0208841P.
PR 07-JUL-2000; 2000US-0216747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR WPI; 2001-616774/71.
DR N-PSDB; AAS59650.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris.
XX
PS Example 1; SEQ ID NO 25958; 1069pp; English.
XX
CC Sequences AAU39105-AAU68017 represent *Propionibacterium acnes* immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC *P. acnes*. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC *P. acnes* is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of *P. acnes* in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for *P. acnes* proteins. These antibodies can be used to
CC downregulate expression and activity of *P. acnes* polypeptides and
CC therefore treat *P. acnes* infections. The antibodies may also be used as
CC diagnostic agents for determining *P. acnes* presence, for example, by
CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX

SQ Sequence 144 AA;
 Query Match 76.5%; Score 39; DB 4; Length 144;
 Best Local Similarity 70.0%; Pred. No. 16;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 SANHSGNASQ 10
 |||||:
 Db 12 SANHNGHGSQ 21

RESULT 4
 ABM61282
 ID ABM61282 standard; protein; 144 AA.
 XX AC ABM61282;
 XX DT 20-OCT-2003 (first entry)
 XX DE Propionibacterium acnes predicted ORF-encoded polypeptide #25958.
 XX KW Acne vulgaris; antiseborrheic; dermatological; antibacterial;
 XX KW immunostimulant; immune response; vaccine.
 XX OS Propionibacterium acnes.
 XX OS WO2003033515-A1.
 XX PN 24-APR-2003.
 XX PD
 XX PF 11-OCT-2002; 2002WO-US032727.
 XX PR 15-OCT-2001; 2001US-00978825.
 XX PR (CORI-) CORIXA CORP.
 XX PA Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;
 XX PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;
 XX PI Barth B, Vallie-Douglass J;
 XX XX WPI: 2003-381789/36.
 XX DR N-PSDB; ACF64579.
 XX PT New Propionibacterium acnes polypeptides and polynucleotides encoding the
 XX PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
 XX PT or for stimulating an immune response specific for a P. acnes protein.
 XX PS Example 1; SEQ ID NO 25958; 1481pp; English.
 XX CC The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
 XX CC encoding a Propionibacterium acnes protein. The invention also relates to
 XX CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
 XX CC immunogenic fragments of P. acnes polypeptides. The invention
 XX CC additionally encompasses expression vectors and host cells comprising a
 XX CC polynucleotide of the invention; antibodies against polypeptides of the
 XX CC invention; fusion proteins comprising a polypeptide of the invention; a
 XX CC method for stimulating an immune response specific for a P. acnes
 XX CC polypeptide and an isolated T cell population comprising T cells prepared
 XX CC via this method; a vaccine composition (comprising P. acnes polypeptides,
 XX CC polynucleotides, antibodies, fusion proteins, T cell populations, or
 XX CC antigen-presenting cells that express the polypeptide); a method and kit
 XX CC for detecting or determining the presence or absence of P. acnes in a
 XX CC patient; and a method for inhibiting the development of P. acnes in a
 XX CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
 XX CC proteins, T cell populations or antigen-presenting cells that express the
 XX CC polypeptides are useful for diagnosing, preventing or treating acne
 XX CC vulgaris, or for stimulating an immune response specific for a P. acnes
 XX CC protein. The polynucleotides can also be used as probes or primers for
 XX CC nucleic acid hybridisation. The vaccine composition is useful for the
 XX CC stimulation of an immune response against P. acnes, or for treating acne,
 XX CC and the kit is useful for performing a diagnostic assay. The present
 XX CC sequence represents a polypeptide predicted to be encoded by an ORF (open
 XX CC reading frame) contained within the P. acnes polynucleotides of the

CC invention. Note: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 144 AA;
 Query Match 76.5%; Score 39; DB 6; Length 144;
 Best Local Similarity 70.0%; Pred. No. 16;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 SANHSGNASQ 10
 |||||:
 Db 12 SANHNGHGSQ 21

RESULT 5
 ABU47415
 ID ABU47415 standard; protein; 1461 AA.
 XX AC ABU47415;
 XX XX 19-JUN-2003 (first entry)
 XX DE Protein encoded by Prokaryotic essential gene #32942.
 XX XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX OS Salmomella typhimurium.
 XX PN WO200277183-A2.
 XX PD 03-OCT-2002.
 XX PF 21-MAR-2002; 2002WO-US009107.
 XX PR 21-MAR-2001; 2001US-00815242.
 XX PR 06-SEP-2001; 2001US-00948993.
 XX PR 25-OCT-2001; 2001US-0342923P.
 XX PR 08-FEB-2002; 2002US-00072851.
 XX PR 06-MAR-2002; 2002US-0362699P.
 XX PA (ELIT-) ELITRA PHARM INC.
 XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX XX WPI: 2003-0299926/02.
 XX DR N-PSDB; ACA51285.
 XX PS Claim 25; SEQ ID NO 75339; 1766pp; English.
 XX CC The invention relates to an isolated nucleic acid comprising any one of
 XX CC the 6213 antisense sequences given in the specification where expression
 XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 XX CC (1) a vector comprising a promoter operably linked to the nucleic acid
 XX CC encoding a polypeptide whose expression is inhibited by the antisense
 XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 XX CC polypeptide or its fragment whose expression is inhibited by the
 XX CC antisense nucleic acid; (4) an antibody capable of specifically binding
 XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 XX CC proliferation or the activity of a gene in an operon required for
 XX CC proliferation; (7) identifying a compound that influences the activity of
 XX CC the gene product or that has an activity against a biological pathway
 XX CC required for proliferation, or that inhibits cellular proliferation; (8)
 XX CC identifying a gene required for cellular proliferation or the biological
 XX CC pathway in which a proliferation-required gene or its gene product lies
 XX CC or a gene on which the test compound that inhibits proliferation of an
 XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 XX CC compound's activity; (11) a culture comprising strains in which the gene

CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 1461 AA;

Query Match 76.5%; Score 39; DB 6; Length 1461;
 Best Local Similarity 70.0%; Pred. No. 2.2e+02;
 Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 SANHSGNASQ 10
 DB 328 SASHTGNASK 337
 ||:|:||||:

RESULT 6
 ABP78403
 ID ABP78403 standard; protein; 283 AA.

XX AC ABP78403;
 XX DT 07-MAR-2003 (first entry)
 XX DE N. gonorrhoeae amino acid sequence SEQ ID 3336.
 XX KW Antibacterial; infection; vaccine; gene therapy.
 XX OS Neisseria gonorrhoeae.
 XX PN WO200279243-A2.
 XX PD 10-OCT-2002.

XX PF 12-FEB-2002; 2002WO-IB002069.

XX PR 12-FEB-2001; 2001GB-00003424.

XX PA (CHIR-) CHIRON SPA.

XX PI Fontana MR, Pizza M, Massignani V, Monaci E;

XX DR WPI; 2003-058415/05.

XX DR N-PSDB; AB239373.

XX PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a
 PT medicament for treating or preventing N. gonorrhoeae infection.

XX PS Disclosure; Page 433; 815pp; English.

XX CC The present invention relates to proteins from Neisseria gonorrhoeae.
 CC Also disclosed are the nucleic acid molecules encoding the proteins and
 CC antibodies that specifically bind to the proteins. The composition
 CC comprising the protein, nucleic acid or antibody is useful for the
 CC manufacture of a medicament for treating or preventing N. gonorrhoeae
 CC infection, this may be in the form of a vaccine or gene therapy.
 CC Sequences given in records ABP76736-ABP81046 represent nucleic acid
 CC molecules of the invention
 XX

SQ Sequence 283 AA;

Query Match 70.6%; Score 36; DB 6; Length 283;
 Best Local Similarity 87.5%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ANHSGNAS 9
 DB 136 AEHSGNAS 143
 |||||:

RESULT 7
 ABB60992
 ID ABB60992 standard; protein; 1186 AA.

XX AC ABB60992;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 9768.

XX KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US009231.

XX PR 23-MAR-2000; 2000US-0191637P.

XX PR 11-JUL-2000; 2000US-00614150.

XX PA (PEKE) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI; 2001-656860/75.

XX DR N-PSDB; ABL05095.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.

XX PS Disclosure; SEQ ID NO 9768; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX

SQ Sequence 1186 AA;

Query Match 70.6%; Score 36; DB 4; Length 1186;
 Best Local Similarity 75.0%; Pred. No. 6.8e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 NHSGNASQ 10
 DB 397 NHSGNAQE 404
 |||||:

RESULT 8
 ABP78234
 ID ABP78234 standard; protein; 328 AA.

XX AC ABP78234;

XX DT 07-MAR-2003 (first entry)

XX

DE N. gonorrhoeae amino acid sequence SEQ ID 2998.
 KW Antibacterial; infection; vaccine; gene therapy.
 XX
 OS Neisseria gonorrhoeae.
 XX
 PN WO200279243-A2.
 XX
 PD 10-OCT-2002.
 XX
 PF 12-FEB-2002; 2002WO-IB002069.
 XX
 PR 12-FEB-2001; 2001GB-00003424.
 XX
 PA (CHIR-) CHIRON SPA.
 XX
 PI Fontana MR, Pizza M, Masignani V, Monaci E;
 XX
 DR WPI; 2003-058415/05.
 XX
 DR N-PSDB; AB239204.
 XX
 PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a
 PT medicament for treating or preventing N. gonorrhoeae infection.
 XX
 PS Disclosure; Page 407; 815pp; English.
 XX
 CC The present invention relates to proteins from Neisseria gonorrhoeae.
 CC Also disclosed are the nucleic acid molecules encoding the proteins and
 CC antibodies that specifically bind to the proteins. The composition
 CC comprising the protein, nucleic acid or antibody is useful for the
 CC manufacture of a medicament for treating or preventing N. gonorrhoeae
 CC infection, this may be in the form of a vaccine or gene therapy.
 CC Sequences given in records ABP76736-ABP81046 represent nucleic acid
 CC molecules of the invention
 XX
 SQ Sequence 328 AA;
 Query Match 68.8%; Score 35; DB 6; Length 328;
 Best Local Similarity 75.0%; Pred. No. 2.5e+02;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 3 NHSGNASQ 10
 DB 75 NHSGNVGQ 82
 |||||
 AAR91515
 ID AAR91515 standard; protein; 341 AA.
 AC AAR91515;
 XX
 DT 16-OCT-2003 (revised)
 DT 27-AUG-2003 (revised)
 DT 25-MAR-2003 (revised)
 DT 25-SEP-1996 (first entry)
 DE Listeria phage lysin PLY511.
 XX
 KW Lysin; phage lysin; PLY; detection; bacteria; Listeria.
 XX
 OS Listeria monocytogenes; phage A511.
 XX
 PN WO9607756-A1.
 PD 14-MAR-1996.
 XX
 PF 07-SEP-1995; 95WO-EP003512.
 XX
 PR 09-SEP-1994; 94DE-04432053.
 PR 24-FEB-1995; 95DE-01006615.
 PR 06-MAR-1995; 95EP-00301416.
 XX

(MERE) MERCK PATENT GMBH.
 Scherer S, Loessner M, Stewart GS, Schubert P;
 WPI; 1996-171627/17.
 N-PSDB; AAT13340.
 Detection of bacteria after lysis with phage lysin enzymes - including
 new Listeria phage lysin(s).
 Claim 13; Page 16-17; 27pp; German.
 Detection of bacteria is effected by lysing bacterial cells with a
 purified phage lysin (PLY) and detecting bacterial cell components in the
 lysate. Target analytes (e.g. ATP, chromosomal DNA or ribosomal RNA) can
 be released in high yields under mild conditions without the safety
 problems associated with intact bacteriophage (cf. WO9406931). Three new
 Listeria phage lysins PLY118, PLY500 and PLY511 are provided (AAT13338-
 T13340). The new PLYs can be isolated from Listeria spp. infected with
 phage A118, A500 or A511, or can be produced by recombinant DNA
 techniques, e.g. using E. coli strains transformed with expression
 plasmids contg. the ply118, ply500 or ply511 gene. (Updated on 25-MAR-
 2003 to correct PR field.) (Updated on 27-AUG-2003 to correct OS field.)
 (Updated on 16-OCT-2003 to standardise OS field)
 Sequence 341 AA;
 Query Match 68.6%; Score 35; DB 2; Length 341;
 Best Local Similarity 77.8%; Pred. No. 2.6e+02;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 2 ANHSGNASQ 10
 DB 231 ANYSGTASQ 239
 ||:|||||
 RESULT 10
 ABU34479
 ID ABU34479 standard; protein; 418 AA.
 XX
 AC ABU34479;
 XX
 DT 19-JUN-2003 (first entry)
 XX
 DE Protein encoded by Prokaryotic essential gene #20006.
 XX
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX
 OS Mycobacterium bovis.
 XX
 PN WO200277183-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 21-MAR-2002; 2002WO-US009107.
 XX
 PR 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX
 DR WPI; 2003-029926/02.
 DR N-PSDB; ACA38349.
 XX
 PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.

XX PS Claim 25; SEQ ID NO 62403; 1766pp; English.

XX CC The invention relates to an isolated nucleic acid comprising any one of

XX CC the 6213 antisense sequences given in the specification where expression

XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:

XX CC (1) a vector comprising a promoter operably linked to the nucleic acid

XX CC encoding a polypeptide whose expression is inhibited by the antisense

XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated

XX CC polypeptide or its fragment whose expression is inhibited by the

XX CC antisense nucleic acid; (4) an antibody capable of specifically binding

XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

XX CC proliferation or the activity of a gene in an operon required for

XX CC proliferation; (7) identifying a compound that influences the activity of

XX CC the gene product or that has an activity against a biological pathway

XX CC required for proliferation, or that inhibits cellular proliferation; (8)

XX CC identifying a gene required for cellular proliferation or the biological

XX CC pathway in which a proliferation-required gene or its gene product lies

XX CC or a gene on which the test compound that inhibits proliferation of an

XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a

XX CC compound's activity; (11) a culture comprising strains in which the gene

XX CC product is overexpressed or underexpressed; (12) determining the extent

XX CC to which each of the strains is present in a culture or collection of

XX CC strains; or (13) identifying the target of a compound that inhibits the

XX CC proliferation of an organism. The antisense nucleic acids are useful for

XX CC identifying proteins or screening for homologous nucleic acids required

XX CC for cellular proliferation to isolate candidate molecules for rational

XX CC drug discovery programs, or for screening homologous nucleic acids

XX CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,

XX CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of

XX CC the target prokaryotic essential genes. Note: The sequence data for this

XX CC patent did not form part of the printed specification, but was obtained

XX CC in electronic format directly from WIPO at

XX CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 418 AA;

Query Match 68.6%; Score 35; DB 6; Length 418;

Best Local Similarity 85.7%; Pred. No. 3.2e+02;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 NHSGNAS 9

Db 146 NHAGNAS 152

RESULT 11

ID ABU36792

XX ABU36792 standard; protein; 418 AA.

XX AC ABU36792;

XX DT 19-JUN-2003 (first entry)

XX DE Protein encoded by Prokaryotic essential gene #22319.

XX DE Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX KW Mycobacterium tuberculosis.

XX OS WO200277183-A2.

XX PN 03-OCT-2002.

XX PD 21-MAR-2002; 2002WO-US009107.

XX PF 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362699P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KU, Zyskind JW;

XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX DR WPI; 2003-029926/02.

XX DR N-PSDB; ACA0662.

XX PT New antisense nucleic acids, useful for identifying proteins or screening

XX PT for homologous nucleic acids required for cellular proliferation to

XX PT isolate candidate molecules for rational drug discovery programs.

XX PS Claim 25; SEQ ID NO 64716; 1766pp; English.

XX CC The invention relates to an isolated nucleic acid comprising any one of

XX CC the 6213 antisense sequences given in the specification where expression

XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:

XX CC (1) a vector comprising a promoter operably linked to the nucleic acid

XX CC encoding a polypeptide whose expression is inhibited by the antisense

XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated

XX CC polypeptide or its fragment whose expression is inhibited by the

XX CC antisense nucleic acid; (4) an antibody capable of specifically binding

XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

XX CC proliferation or the activity of a gene in an operon required for

XX CC proliferation; (7) identifying a compound that influences the activity of

XX CC the gene product or that has an activity against a biological pathway

XX CC required for proliferation, or that inhibits cellular proliferation; (8)

XX CC identifying a gene required for cellular proliferation or the biological

XX CC pathway in which a proliferation-required gene or its gene product lies

XX CC or a gene on which the test compound that inhibits proliferation of an

XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a

XX CC compound's activity; (11) a culture comprising strains in which the gene

XX CC product is overexpressed or underexpressed; (12) determining the extent

XX CC to which each of the strains is present in a culture or collection of

XX CC strains; or (13) identifying the target of a compound that inhibits the

XX CC proliferation of an organism. The antisense nucleic acids are useful for

XX CC identifying proteins or screening for homologous nucleic acids required

XX CC for cellular proliferation to isolate candidate molecules for rational

XX CC drug discovery programs, or for screening homologous nucleic acids

XX CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,

XX CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of

XX CC the target prokaryotic essential genes. Note: The sequence data for this

XX CC patent did not form part of the printed specification, but was obtained

XX CC in electronic format directly from WIPO at

XX CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 418 AA;

Query Match 68.6%; Score 35; DB 6; Length 418;

Best Local Similarity 85.7%; Pred. No. 3.2e+02;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 NHSGNAS 9

Db 146 NHAGNAS 152

RESULT 12

ID ADC94648

XX ADC94648 standard; protein; 762 AA.

XX AC ADC94648;

XX DT 01-JAN-2004 (first entry)

XX DE E. faecium protein sequence SEQ ID 4275.

XX KW Vaccine; urinary tract infection; bacteraemia; endocarditis; wound;

XX KW abdominal-pelvic infection.

XX OS Enterococcus faecium.

XX OS US6583275-B1.

XX PN

```
PD 24-JUN-2003.
XX 30-JUN-1998; 98US-00107532.
XX 02-JUL-1997; 97US-0051571P.
XX 14-MAY-1998; 98US-0085598P.
XX (GENO-) GENOME THERAPEUTICS CORP.
PA Doucette-Stamm LA, Bush D;
XX WPI; 2003-799836/75.
XX N-PSDB; ADC90994.
XX New isolated nucleic acid derived from Enterococcus faecium encoding an
PT Enterococcus faecium polypeptide useful for detection, prevention and
PT treatment of a pathological condition resulting from a bacterial
PT infection.
XX
XX Example 1; SEQ ID NO 4275; 243pp; English.
XX
XX The invention relates to an isolated nucleic acid derived from
CC Enterococcus faecium encoding an Enterococcus faecium polypeptide having
CC one of 10 fully defined sequences given in the (or comprising 40
CC sequential nucleotides chosen from any of the nucleic acids, its
CC complement or sequences hybridising to it). Also included are a
CC recombinant vector comprising the nucleic acid operably linked to
CC transcription regulatory element, a cell comprising the vector and a
CC single-stranded probe comprising the nucleic acid. The nucleic acids are
CC chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.
CC The nucleic acid is useful for diagnosing pathological conditions
CC resulting from E. faecium bacterial infection (e.g. urinary tract
CC infection, bacteraemia, endocarditis, wounds and abdominal-pelvic
CC infection) and for screening drugs such as agonists and antagonists. The
CC nucleic acid is useful for recombinant production of Candida albicans -
CC derived peptides or antisense polypeptides. Pharmaceutical compositions
CC and vaccines containing the nucleic acid are useful for preventing or
CC treating Enterococcus faecium infections. The present sequence represents
CC one if the disclosed E. faecium proteins.
XX
XX Sequence 762 AA;
XX
XX Query Match 68.6%; Score 35; DB 7; Length 762;
XX Best Local Similarity 85.7%; Pred. No. 6.4e+02;
XX Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 SANHSGN 7
XX |:|||||
XX 43 SSNHSGN 49
XX
XX RESULT 13
XX ABU21849
XX ID ABU21849 standard; protein; 1246 AA.
XX AC
XX ABU21849;
XX
XX 19-JUN-2003 (first entry)
XX
XX Protein encoded by Prokaryotic essential gene #7376.
XX
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX Burkholderia fungorum.
XX
XX WO200277183-A2.
XX
XX 03-OCT-2002.
XX
XX 21-MAR-2002; 2002WO-US009107.
XX
XX 21-MAR-2001; 2001US-00815242.
XX
XX 06-SEP-2001; 2001US-00948993.
XX
XX 25-OCT-2001; 2001US-0342923P.
XX 08-FEB-2002; 2002US-00072851.
XX 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
XX WPI; 2003-029926/02.
XX N-PSDB; ACA25719.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 25; SEQ ID NO 49773; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 1246 AA;
XX
XX Query Match 68.6%; Score 35; DB 6; Length 1246;
XX Best Local Similarity 70.0%; Pred. No. 1.1e+03;
XX Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 1 SANHSGN 10
XX |:|||||
XX 1090 SVNSGNAQ 1099
XX
XX RESULT 14
XX ABB52677
XX ID ABB52677 standard; protein; 1778 AA.
XX
XX ABB52677;
XX
XX 11-FEB-2002 (first entry)
XX
XX Escherichia coli polypeptide SEQ ID NO 749.
XX
XX Escherichia coli; B2/D+A-; antinflammatory; antibacterial;
XX immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
```

KW systemic infection; non-diarrhoeal infection; septicaemia;
 KW pyelonephritis; antibiotic resistance.

OS Escherichia coli.

XX WO200166572-A2.

XX 13-SEP-2001.

XX 12-MAR-2001; 2001WO-EP003445.

XX 10-MAR-2000; 2000FR-00003145.

PR 02-FEB-2001; 2001FR-00001449.

XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

PI Bingen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;

XX WPI; 2001-550253/61.

XX A library of DNA fragments of Escherichia coli strains for the phylogenetic
 PT determination of a given strain comprises polynucleotides of nature B2/D+
 PT A-.

PS Example 6; Fig 6; 646pp; English.

CC The invention relates to a library of DNA fragments of Escherichia coli
 CC strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533) and
 CC encoded proteins (ABE52459-ABE52919 and ABE52954-ABE53094) of nature
 CC B2/D+A-. The polynucleotides have potential anti-inflammatory,
 CC antibacterial and immunosuppressive activity as part of pharmaceutical
 CC compositions used to treat, palliate or prevent extra-intestinal E. coli
 CC infections. The polypeptides are useful for determining the phylogenetic
 CC group of a given E. coli strain. These polypeptides can detect and treat
 CC an undesired development of E. coli, particularly an extra-intestinal
 CC infection that include systemic and non-diarrhoeal infections such as
 CC septicaemia, pyelonephritis and meningitis this is particularly
 CC advantageous as bacterial resistance is increasing with the more frequent
 CC use of broad spectrum antibiotics

XX Sequence 1778 AA;

Query Match 68.6%; Score 35; DB 4; Length 1778;

Best Local Similarity 60.0%; Pred. No. 1.7e+03;

Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 SANHSGNASQ 10

Db :||| |||||

1180 NASHDGNASK 1189

RESULT 15

ABP70279
 ID ABP70279 standard; peptide; 12 AA.

XX AC ABP70279;

XX DT 07-APR-2003 (first entry)

XX DE Amino acid sequence of a mimetic of the core protein of HCV.

XX KW Antigen; core protein; HCV; antibody; immunogen; vaccine.

XX OS Synthetic.

XX OS Hepatitis C virus.

XX PN WO200296929-A2.

XX PD 05-DEC-2002.

XX PF 31-MAY-2002; 2002WO-FR001851.

XX PR 31-MAY-2001; 2001FR-00007184.

XX (INRM) BIOMERIEUX SA.

XX Jolivet Reynaud C;

XX WPI; 2003-140440/13.

XX New polypeptides specifically reactive with hepatitis C virus antibodies,
 PT useful for diagnosis of, and in vaccines against hepatitis C.

XX Claim 1; Page 14; 36pp; French.

XX The present sequence represents a mimetic peptide, derived from the core
 CC protein of Hepatitis C virus (HCV). The specification describes peptides
 CC that react specifically with antibodies in patients infected with
 CC Hepatitis C virus (HCV). The peptides are used as reagents for detection
 CC and quantification of anti-HCV antibodies in a patient. They are also
 CC used for the preparation of specific antibodies, which are useful for
 CC detection and quantification of HCV core protein. The peptides are also
 CC used as immunogens in immunotherapeutic compositions, particularly
 CC vaccines

XX Sequence 12 AA;

Query Match 66.7%; Score 34; DB 6; Length 12;

Best Local Similarity 75.0%; Pred. No. 9.1;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 3 NHSGNASQ 10

Db :||| |||||
 5 SHSGNAKQ 12

Search completed: July 18, 2005, 13:40:32

Job time : 87.2 secs